

UNIVERSITY OF HAWAII
LIBRARY
SEP 3 '60

The Philosophical Magazine

FIRST PUBLISHED IN 1798

A Journal of Theoretical Experimental and Applied Physics

Vol. 5

June 1960
Eighth Series

No. 54

25s. 0d., plus postage
Annual Subscription £13 10s. 0d., payable in advance



Printed and Published by

TAYLOR & FRANCIS LTD
RED LION COURT, FLEET STREET, LONDON, E.C.4

THE PHILOSOPHICAL MAGAZINE

Editor

Professor N. F. MOTT, M.A., D.Sc., F.R.S.

Editorial Board

Sir LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.

Sir GEORGE THOMSON, M.A., D.Sc., F.R.S.

Professor A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

AUTHORS wishing to submit papers for publication in the Journal should send manuscripts directly to the Publishers.

Manuscripts should be typed in *double spacing* on one side of quarto (8×10 in.) paper, and authors are urged to aim at absolute clarity of meaning and an attractive presentation of their texts.

References should be listed at the end in alphabetical order of authors and should be cited in the text in terms of author's name and date. Diagrams should normally be in Indian ink on white card, with lettering in soft pencil, the captions being typed on a separate sheet.

A leaflet giving detailed instructions to authors on the preparation of papers is available on request from the Publishers.

Authors are entitled to receive 25 offprints of a paper in the Journal free of charge, and additional offprints can be obtained from the Publishers.

The *Philosophical Magazine* and its companion journal, *Advances in Physics*, will accept papers for publication in experimental and theoretical physics. The *Philosophical Magazine* publishes contributions describing new results, letters to the editor and book reviews. *Advances in Physics* publishes articles surveying the present state of knowledge in any branch of the science in which recent progress has been made. The editors welcome contributions from overseas as well as from the United Kingdom, and papers may be published in English, French and German.

Nuclear Interactions of 5.7 BeV Protons in Photographic Emulsion†

By V. Y. RAJOPADHYE

H. H. Wills Physical Laboratory, University of Bristol‡

[Received October 1, 1959]

ABSTRACT

Tracks produced by 5.7 bev protons in nuclear emulsions have been followed till they interact with emulsion nuclei, and a mean free path of 35.6 ± 2.4 cm has been obtained. About 24% of the shower tracks are found to be protons, all of which emerge at angles less than 20° with the direction of the primary. The mean multiplicity of pion production is found to be 2.6 ± 0.3 , the coefficient of inelasticity 0.49 ± 0.07 , and the transverse momentum of pions 220 ± 30 mev/c.

The angular distribution of the shower tracks has been analysed on the assumption that the pions are produced in a single collision system with a forward-backward symmetry in the centre of mass reference frame. The mass of the particle with which the incident proton forms such a collision, is shown to have an upper limit of 1.08 ± 0.15 proton mass units. Hence the meson production on an average may be considered as arising in a nucleon-nucleon collision even with a complex target nucleus. The angular distribution of the pions in the C-system is found to be isotropic, within the limits of experimental errors.

§ 1. INTRODUCTION

HIGH energy protons and α -particles from cosmic radiation have provided extensive data on multiple meson production. Such observations have been made both with cloud chambers and photographic emulsions. In those cases where few or no heavily ionizing prongs are associated with the event, it has been customary to assume that the secondary particles are created in a single collision of the incident proton with an individual nucleon of the target nucleus. The energy of the incident proton is then estimated from the median angle of the shower tracks, i.e. the angle which contains half the number of shower tracks (Dilworth *et al.* 1953).

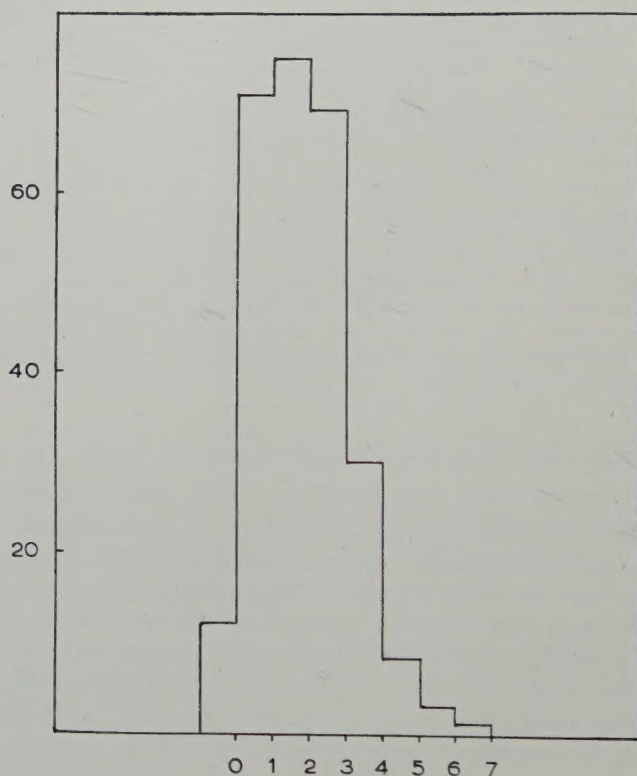
During recent years, beams of high energy protons have become available, and diffusion chambers and bubble chambers filled with liquid hydrogen have been used to investigate proton-proton interactions at various energies. These energies, though small as compared with those available in the cosmic radiation, are much above the threshold for multiple meson production. While photographic emulsion is not the ideal medium for studying such nucleon-nucleon collisions, it has nevertheless often been used for this

† Communicated by the Author.

‡ On leave of absence from the S. P. College, Poona, India.

purpose (Duke *et al.* 1955, Cester *et al.* 1956, Kalbach *et al.* 1959). Moreover, proton-nucleus collisions in the emulsion are capable of providing a wealth of useful information. It is intended to give here an analysis of such collisions, initiated by 5.7 Bev protons, in relation to various aspects of meson production, such as multiplicity, angular distribution, inelasticity, etc. Although heavy mesons, hyperons and anti-nucleons, can be produced at this energy, the cross section for their production is very much smaller than for π mesons; hence it is assumed that all the newly created particles are π mesons. The primary energy being known, the angular distribution has been used to examine the hypothesis of nucleon-nucleon collision.

Fig. 1



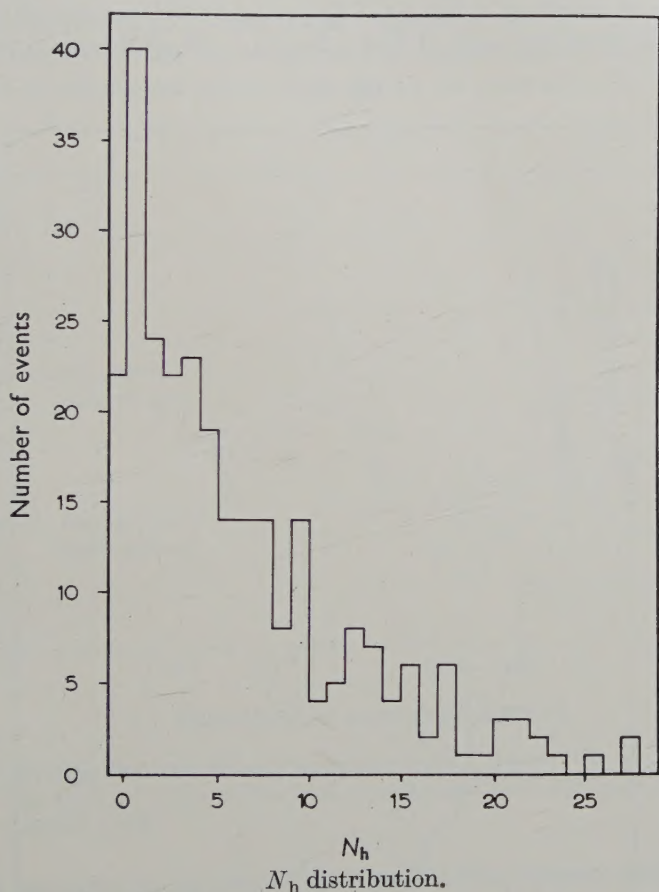
n_s distribution.

§ 2. EXPERIMENTAL

A small stack of Ilford G5 emulsions ($10\text{ cm} \times 4\text{ cm} \times 400\mu$) was exposed to a single pulse of the 5.7 Bev proton beam from the Berkeley bevatron. The tracks were parallel to the shorter edge and had negligible dip. The beam had a high flux density of $\sim 5 \times 10^5$ particles per cm^2 . Only two plates of the stack were used in this experiment.

Adopting a procedure of scanning along the track, 266 interactions have been obtained. An interaction has been defined as an event where the beam track seems to terminate, with or without the association of secondary track(s) of any degree of ionization. Events that look like a mere deflection of the incident track were ignored, as they would require a detailed examination, which is often inconclusive, to decide between meson production and elastic scatter. Each interaction was classified according to the number of N_h of the heavily ionizing prongs (black and grey) and the number, n_s , of the shower tracks, i.e. tracks with a blob density less than 1.4 times that of the beam proton tracks. The projected and dip angles of each shower track were measured and the space angle calculated.

Fig. 2



§ 3. PRELIMINARY RESULTS

3.1. n_s and N_h Distributions

The n_s distribution is shown in fig. 1. Our mean value of n_s is 2.3, as compared with 2.1, of Schein *et al.* (1956). Events in which n_s is 1, 2 or 3 are equally abundant and these together constitute 60% of the total.

The N_h distribution is shown in fig. 2. There is a high proportion (viz. 23%) of events in which $N_h \leq 1$. Half the number of events are included in the interval $N_h \leq 4$.

The observations do not show any correlation between N_h and n_s . The mean multiplicity n_s does not seem to vary with N_h , as may be seen from the table below.

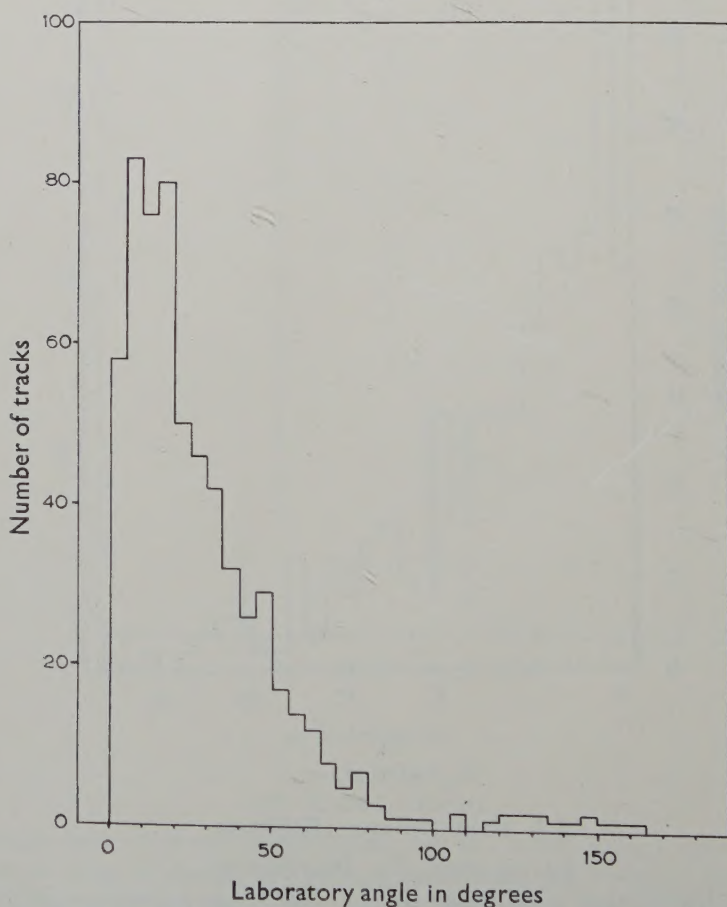
Table 1. Variation of n_s with N_h

N_h	0, 1	2, 3	4-6	7-10	> 11
Mean n_s	2.4	2.2	2.5	2.3	2.2

3.2. Mean Free Path

A total track length of 75.9 m was followed and 213 interactions obtained, thus giving a mean free path of 35.6 ± 2.4 cm. (53 additional interactions

Fig. 3



Differential angular distribution of all the shower tracks of the composite star.

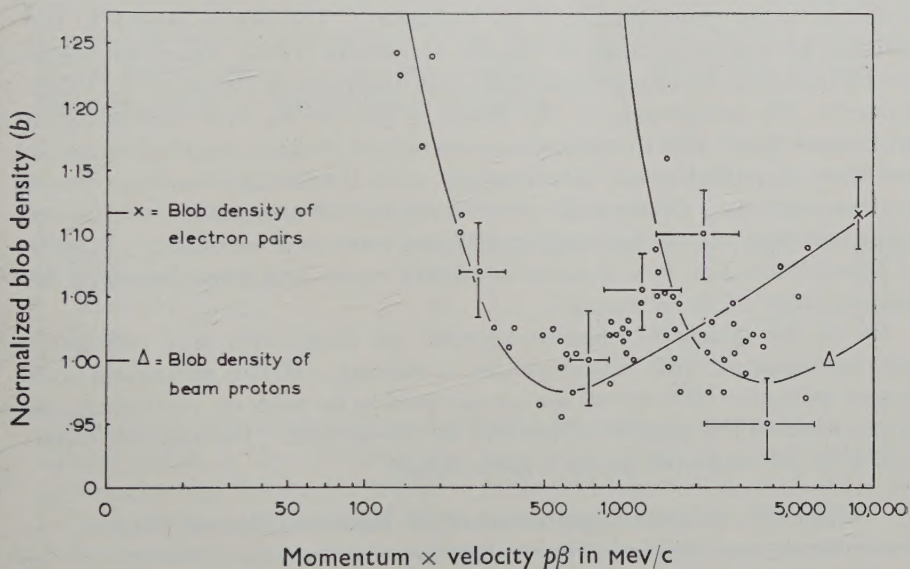
were found, but the path length for these was not recorded. All the 266 interactions have been used for the further analysis.) The mean free path is in good agreement with other observations at the same energy. Cavanaugh *et al.* (1955), obtain 37.6 ± 5.3 cm from 50 events and the Bombay group (Appa Rao *et al.* 1956) find 34.7 ± 3.4 cm from 103 events.

§ 4. MEDIAN ANGLE OF THE π MESONS

Since the average number of shower tracks per star is small, all the events have been grouped together to form a 'composite star', to which it is possible to apply statistical considerations. The angular distribution of this composite star is shown in fig. 3.

Measurements on the blob density, b , and multiple Coulomb scattering, $\bar{\alpha}$, were made on all secondary tracks longer than 5 mm with a view to identifying the particles producing them. The blob density was normalized to that of nearby proton tracks at the appropriate depth. 1000 blobs were counted on the shower tracks and 3000 on the proton tracks.

Fig. 4



Curves showing the blob density *vs.* momentum \times velocity for protons and pions. The plateau value of the blob density has been obtained from tracks of electron pairs.

It is necessary to ascertain that the scattering measurements even after correction for the stage noise and reading errors, are not seriously affected by the existence of distortion, which commonly occurs on account of the differential shear of the emulsion during processing. The effect of such distortion is to impart a curvature to straight tracks. Observations were made on a number of steep grey tracks to determine the magnitude of this

effect in several regions. It was then concluded that the curvature due to this cause, imposed on tracks longer than 5 mm, would introduce an error in the scattering measurement much less than that due to statistical fluctuations arising from the use of a finite number of independent cells.

The second effect to be considered is the spurious scattering (Biswas *et al.* 1955) which becomes an important source of noise in measurements using large cell sizes. Spurious scattering is believed to be caused by local dislocations in the emulsion, completely unrelated to the differential shear mentioned above. From scattering measurements on a large number of beam proton tracks it was found that the spurious scattering in these plates was fairly small, and thus the scattering measurements may be considered to give a satisfactory estimate of the energy of the particle, subject only to the statistical errors associated with the measurements.

The results of the measurements of blob density and $p\beta$ from scattering on 64 tracks which satisfied the length criterion are shown in fig. 4. The plateau value of the blob density was determined from the tracks of electron pairs materialising from the γ -rays created in π^0 -decay. The plateau value of ionization and the shape of the ionization curve are believed to be sensitive to the development of the emulsion. The curves shown in the diagram are similar to those of Shapiro and Stiller (1952), where the rise in the blob density is more gradual than that obtained by Daniel *et al.* (1952). However, the uncertainty in the shape of the curves, and the statistical errors associated with the measurements of blob density and scattering, do not allow an unambiguous identification of all the tracks into those due to protons or pions. Cester *et al.* (1956) have arrived at a similar conclusion, while investigating the interactions of 3 BeV protons in emulsion. In spite of these difficulties, it is possible to obtain upper and lower limits to the median angle of the π mesons.

(a) In the sample of measured tracks, all those with $p\beta \leq 1200$ meV/c can be classified with confidence as π mesons. If the remaining, (i.e. tracks with $p\beta > 1200$ meV/c) are all assumed to be protons, one obtains an upper limit to the number of protons in the sample. The angular distribution of the measured tracks is given below.

Table 2. Angular Distribution of the Measured Shower Tracks

Angle	0-10°	10°-20°	20°-30°	>30°	Total
Number of tracks with $p\beta > 1200$ meV/c	32	2	0	0	34
Total number of tracks	43	6	4	11	64

Assuming this upper limit to the fraction of protons also holds in the corresponding angular intervals of all the shower tracks (of the composite star), one obtains an upper limit to the number of protons as 157^\dagger . The

† Total number of shower particles is 606.

median angle of the remaining, considered to be π mesons, would then be $28.9^\circ \pm 1.6^\circ$, since none of the tracks that could be ascribed to protons, are observed at angles greater than 20° . This median angle is an upper limit to the true median angle of the pions alone, since under the assumption, the number of pions at small angles is underestimated.

(b) If all the shower tracks of the composite star are considered to be π mesons, the median angle would be $20.7^\circ \pm 0.8^\circ$. This represents a lower limit to the true median angle of the pions since the number of pions at small angles is over estimated in this case.

An independent estimate of the median angle may be obtained in the following way:

(c) From statistical considerations, the incident proton will re-emerge as a proton in about half the number of collisions, and as a neutron in the other half. If in all cases, the secondary protons are relativistic, approximately 133 shower tracks emitted at small angles could be ascribed to such protons. The median angle of the remaining, all assumed to be pions, would be $27.7^\circ \pm 1.6^\circ$.

(d) Notwithstanding the remarks made above, about the use of (b^* , $p\beta$) curves, it would be interesting to see what proportions of protons among the shower tracks is predicted if the tracks are identified (as protons or pions) with their aid in the best possible way. When this is done, the following angular distribution is obtained.

Table 3

Interval	0-10°	10°-20°	20°-30°	> 30°	Total
Number of protons	21	2	0	0	23
Number of pions	22	4	4	11	41
Total	43	6	4	11	64

Assuming the same proportions of protons and pions to hold for all the shower tracks in the corresponding angular intervals, the total number of protons among the shower tracks is estimated (all of these are within 20°) to be 121; the median angle of the pions would then be $26.9^\circ \pm 1.6^\circ$.

The median angle in the neighbourhood of this value does not seem to be very sensitive to the errors in the estimate of the number of protons. In fact the results under assumptions (a), (c) and (d) are equal within limits of experimental errors.

§ 5. KINEMATICS OF MESON PRODUCTION

Let it be assumed that the incident proton, colliding with a particle or group of particles within the target nucleus (hereafter designated as the collision partner) produces all the mesons in a single encounter. We can then define the C-system, i.e. the coordinate system in which the centre

of mass of the colliding particles is at rest. This system has a velocity β_c (in units of velocity of light) with respect to the laboratory system in which the target nucleus is at rest. β_c is given by the relation,

$$\beta_c = \frac{pc}{E + E_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where p and E are the momentum and total energy of the incident proton and E_2 the rest energy of the collision partner.

A meson emitted with a velocity β^* in the C-system at an angle ϕ to the shower axis, i.e. the direction of motion of the incident proton, will appear in the laboratory system to be ejected at an angle θ . The relation between θ and ϕ is

$$\gamma_c \tan \theta = \frac{\sin \phi}{\beta_c / \beta^* + \cos \phi} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where

$$\gamma_c = 1 / \sqrt{1 - \beta_c^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2a)$$

Assuming forward-backward symmetry in the C-system, half the mesons would be emitted within an angle η in the laboratory system (which corresponds to $\pi/2$ in the C-system), given by

$$\gamma_c \beta_c \tan \eta = \beta^*. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The total energy of the meson in the C-system, γ^* , expressed in terms of the energy (γ) and momentum ($\beta\gamma$) in the L-system provides another useful relation, viz:

$$\gamma^* = \gamma \gamma_c (1 - \beta_c \beta \cos \theta) \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where $\gamma\beta_x = \gamma\beta \cos \theta$ is the component along the shower axis, of the momentum of the meson in the L-system.

The measurements of the angles of emission, and multiple Coulomb scattering of the identified pions (see table 3) provide the values of γ , β and β_x . It will be recalled that the tracks available for the measurements are biased in favour of those emitted at small angles, because of the selection criterion. The geometrical loss for large angles has to be corrected for by means of an appropriate weight factor, and the true (weighted) mean values of γ , β and β_x determined.

It has been found that these values are very little affected by the errors of identification of the shower tracks. The ambiguity of identification occurs only for tracks of high energy and these are mostly emitted at small angles. The more serious errors in the mean values of β and β_x arise mostly due to the small numbers of tracks at wide angles, and the corresponding large weight factors, apart from the large statistical errors in the measurement of Coulomb scattering.

Equations (3) and (4) may now be solved for the two unknowns, β_c and β^* with the appropriate value of the median angle η . For this purpose it was found convenient to start with an assumed value of β_c (chosen appropriate to a proton-nucleon collision) and work by the method of successive approximations. The mass of the collision partner is then obtained by substituting in (1) the final convergent value of β_c .

From the lower and upper limits of the median angle (viz. 20.7° and 28.9°) the limits for the mass of the collision partner M in units of proton mass are found to be:

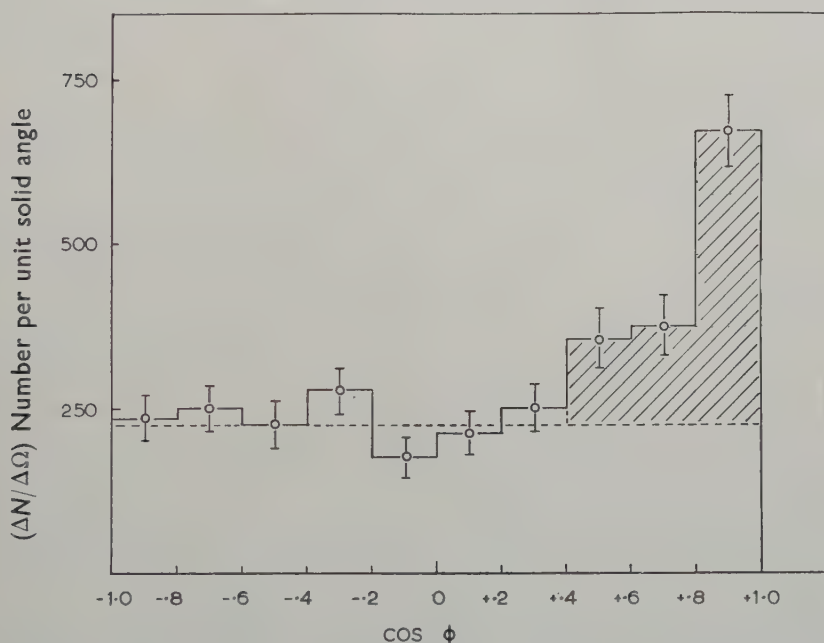
$$\begin{aligned} M_{\min} &= 0.46 \pm 0.05, \\ M_{\max} &= 1.08 \pm 0.15. \end{aligned}$$

The median angles obtained under assumptions (c) and (d) give for M the values of 0.98 ± 0.15 and 0.91 ± 0.15 proton mass units respectively. Hence one may conclude that the meson production, on an average, occurs as if the incident proton suffers a collision with a particle of protonic mass, within the complex target nucleus.

§ 6. ISOTROPY OF EMISSION

Conversion from the measured angles to the C-system angles may now be done by means of the transformation eqn. (2) of the previous section, with β_c , γ_c chosen appropriate to a proton-nucleon collision. β^* may be obtained in several ways, e.g. using eqn. (3), and the median angle under assumption (c) which, as we have seen earlier, gives a value of M very nearly equal to unity.

Fig. 5



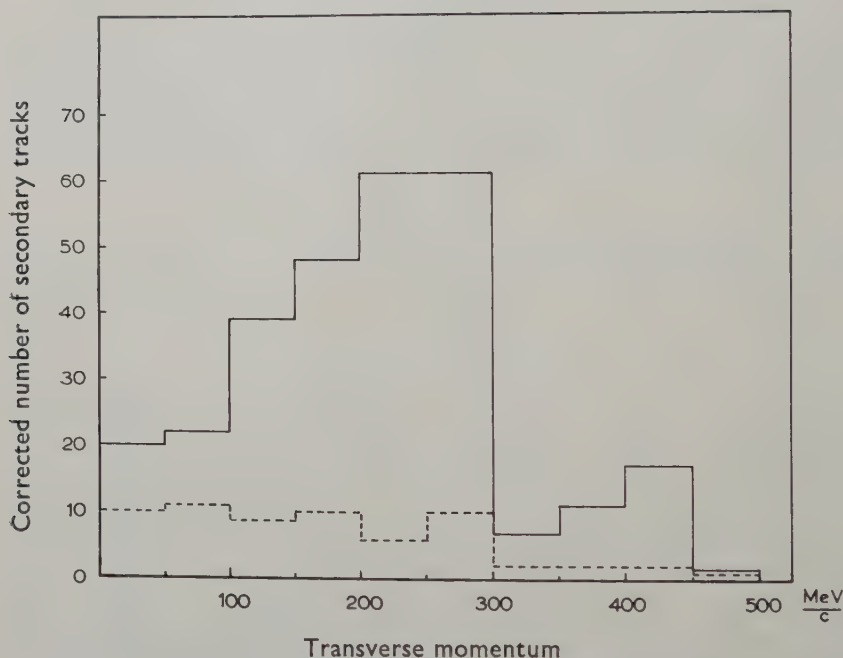
C-system angular distribution. The shaded portion represents the contribution due to protons. The number of these is estimated to be 145 ± 20 .

This allows the C-system angular distribution to be determined for all the shower tracks of the composite star. This is shown in fig. 5. When the peak in the forward direction is interpreted as due to fast protons, the

isotropic character of the distribution for the pions becomes immediately obvious. The number of protons in the forward peak is then estimated to 145 ± 20 , all of which are confined to the interval $1 \geq \cos \phi \geq 0.4$, which corresponds to $0-19^\circ$ in the laboratory system. Both these conclusions are consistent with the observations made earlier.

The values of β^* obtained in various ways do not seem to differ from one another by more than a few per cent. To allow for a possible small uncertainty in the value of β^* , the C-system angular distribution has also been determined for two more values of β_c/β^* , viz. 1 and 0.92, and in all cases it is found to be closely similar. Thus it may be concluded that isotropy obtained is a genuine effect.

Fig. 6



P_1 histogram for all shower tracks, corrected for the geometric loss. The dotted line represents the actual number of tracks measured.

Isotropy of emission for pions produced in 6 Bev proton interactions has been obtained by King (1958) and Kalbach *et al.* (1959). However, Daniel *et al.* (1959) have obtained a forward-backward peaking, an effect probably due to their small statistics, involving the use of only those tracks, which were identified as mesons by $(b^*, p\beta)$ curves, and the large weight factors to correct the geometrical loss.

§ 7. TRANSVERSE MOMENTUM

The importance of this dynamical quantity for the understanding of high energy nuclear interactions was first realized by Nishimura (Koba 1956). Extensive measurements on the transverse momentum P_1 , on the

charged and neutral pions emitted from jets were made by Edwards *et al.* (1958). They find P_1 to be 500 mev/c, over a very wide range of primary energies. Recent observations by the Japanese Emulsion Group (1959) on the P_1 of neutral pions in jets with primary energies $10^{12} \div 10^{14}$ ev show a narrow distribution with a peak at about 400 mev/c. They also find no significant variation with either primary energy or pion energy.

In the present experiment, the transverse momentum of all the secondary particles has been determined from the measurements of the Coulomb scattering and the angle of emission. It does not seem to depend upon the identity (π or p) assigned to the particle when $p\beta_1 > 1200$ mev/c and thus it has the same value for all secondary tracks. The histogram of P_1 , for all shower tracks, in which the observations are duly weighted to compensate for the geometrical loss is shown in fig. 6 from which the mean value of P_1 is found to be 220 ± 30 mev/c, somewhat smaller than that observed in 'jets'. It should be noted that a large number of tracks available for measurements are emitted at small angles. Constancy of P_1 would require very high energies for the mesons emitted at small angles. This is clearly impossible in view of the finite energy of the incident proton. With the small statistics available, it is not possible to draw any conclusions regarding the variation, if any, of the transverse momentum with the angles of emission.

§ 8. ENERGY BALANCE

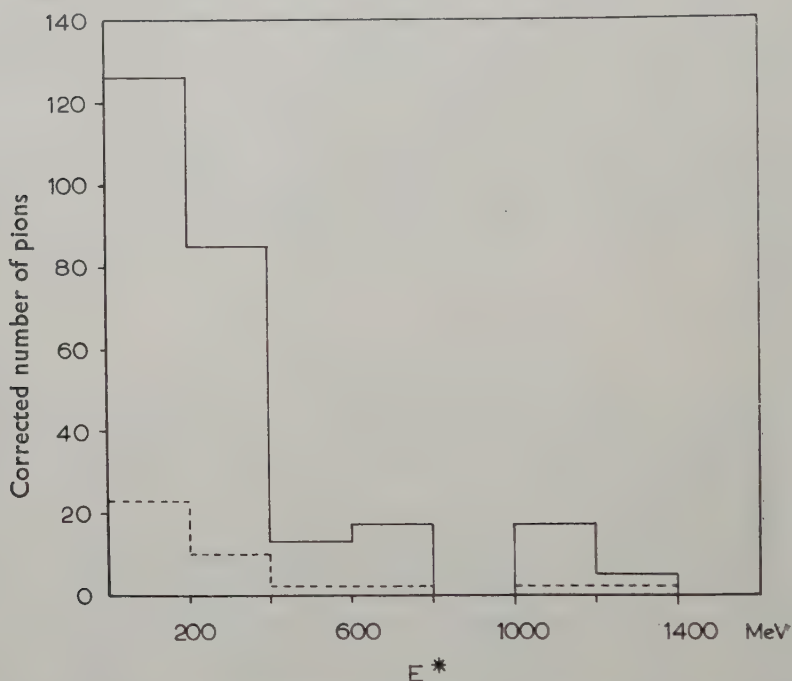
It is possible to see how the kinetic energy of the incident proton is accounted for in the interaction. It is convenient to consider all the interactions together and utilize the results of the scattering measurements on the sample of 64 tracks.

As a starting point, the average kinetic energy of the protons and the total energy of the pions, in the different angular intervals may be determined under assumption (a), § 4. The relative numbers of protons and pions shown in table 2 then allows the energy of all the shower tracks of the composite star to be calculated. A correction has to be made for the energy carried away by the neutrons, and π^0 mesons emitted. About 66% of the collisions would be with the heavy nuclei Ag and Br, in which the neutron proton ratio is 13:10 and thus collisions with neutrons would be more numerous in that ratio. In the case of light nuclei C, N and O, probability of a collision with a proton or a neutron is the same. From all these considerations, the energy released per star, that accounts only for all shower particles (charged and neutral) is found to be 3.5 ± 0.8 mev.

Thus a lower limit is set for the energy of the shower tracks, since the assumption (a) causes the average energy of both pion and proton groups to be underestimated. Assumption (d), which gives a higher proportion of pions, may now be used as comparison, to calculate the average energy of protons and pions. From this, one obtains the total energy of the shower tracks per star to be 4.1 ± 1.1 Bev.

To this must be added the energy of the black and grey tracks. From a study of 9 BeV proton stars in emulsion Beliakov *et al.* (1958) have found the energy (E') of black and grey tracks in light and heavy nuclei to be respectively 470 ± 70 and 1400 ± 120 mev. Considering the proportion of light and heavy nuclei in emulsion, the average energy per star is ~ 1 BeV. This is in good agreement with that deduced from cosmic ray data (Camerini *et al.* (1951), Powell *et al.* (1959)). The latter give $E' = 124N_h + 30$ mev which corresponds to 860 mev in the present experiment. Thus the total energy released per star agrees well with the primary energy.

Fig. 7



Histogram of the C-system kinetic energy for the mesons, corrected for the geometric loss. The dotted line shows the actual number of tracks.

§ 9. PION MULTIPLICITY AND INELASTICITY OF COLLISION

The number of tracks considered to be due to protons among the 606 shower tracks varies according to the assumptions made, but it seems to lie between fairly definite limits of 121 and 157. The number of mesons then would be uncertain to that extent. Mesons emitted backwards in the C-system would appear slow in the laboratory system, and would produce grey tracks. These have not been considered in the present analysis, as their number is presumably small. Assuming the number of protons to be 145 ± 20 as obtained from fig. 5, the number of charged mesons

per star is found to be 1.7. According to the principle of charge independence in meson production, the ratio of neutral to charged mesons is 1:2. Thus the total pion multiplicity would be 2.6 ± 0.3 . This may be compared with 2.5 as predicted by Fermi's theory.

The inelasticity of collision is defined as the fraction of the available energy in the C-system, that is radiated off as mesons. The average energy of the mesons may be computed in several ways. When combined with the mean multiplicity, this would allow the inelasticity to be determined.

The average velocity of the mesons, β^* , has been found to be 0.91. This corresponds to a kinetic energy 210 mev.

The eqns. (3) and (4) of § 5, when solved for β^* and β_c , with the median angle 26.9° under assumption (d) and 27.7° under assumption (c), give β^* as 0.935 and 0.931 respectively. These correspond to energies of 285 and 250 mev.

On the basis of the identification of mesons made under assumption (d), their energies and momenta have been calculated from the measurements of multiple Coulomb scattering. Since the mesons are assumed to be produced in a nucleon-nucleon collision, their C-system energy can be calculated by means of appropriate Lorentz transformation equations. Figure 7 shows the C-system energy spectrum of the mesons, thus obtained, with due corrections for the geometrical loss. The weighted mean kinetic energy is found to be 280 ± 80 mev.

The errors of identification of the tracks do not seem to affect the mean value obtained above. Using assumption (a), for discriminating pions from protons we get nearly the same spectrum, and the same mean energy. The small weight factors associated with the tracks at small angle do not alter noticeably the spectrum by their inclusion or exclusion.

All these results may not be considered inconsistent with one another, since the statistical errors are rather large. Taking the mean energy as 280 ± 80 mev, the inelasticity would be 0.52 ± 0.15 .

Since the statistical errors in the mean energy derived above are too high, we have preferred to use the results of Kalbach *et al.* (1959) who find the mean energy to be 236 ± 20 mev, from measurements on 143 pion tracks. This gives the inelasticity as 0.49 ± 0.07 .

§ 10. DISCUSSION

The median angle method for the analysis of meson showers is based on the assumption of an on-the-average forward-backward symmetry in the C-system. With events of low multiplicity as in the present experiment, the median angle of the shower tracks would be considerably smaller than that for the mesons only, because the proton often emerges from the interaction with only a small change of direction. The primary energy deduced from the median angle of the shower tracks would then be over-estimated (Haber-Schaim 1956). At 6 bev primary energy, the mean velocity of the mesons β^* , in the C-system is found to be somewhat higher

than the velocity of the C-system itself. Ignoring this would have the effect of under-estimating the primary energy. The error involved, however, is of the same order as that due to the statistical error in the median angle.

The fluctuations in the median angle with N_h observed by Haber-Schaim, (1956) are probably due to the small statistical weight of his data. In the present experiment, the variation of the median angle with N_h is found to be as follows:

Table 4. Variation of the Median Angle of Shower Tracks with N_h

N_h Median angle	0 16.3 ± 2.0	1 14.0 ± 1.7	≤ 4 17.5 ± 1.0	≥ 5 25.0 ± 1.4
-----------------------	---------------------	---------------------	----------------------------	----------------------------

A difference between the median angles for small and large N_h groups has been also found by Beliaikov *et al.* (1958), in a study of 9 Bev proton stars. For collision with protons, light and heavy nuclei they find the median angles for shower tracks to be respectively 18° , 25° , and 28° .

Events with small N_h are generally considered to be good approximations to nucleon-nucleon collisions. The number of such events is not very large in this experiment. However, adopting the same procedure as before, the C-system angular distribution for pions, in these events has been determined. A plot similar to fig. 5 shows isotropy of emission for pions, allows the number of protons in the forward peak and hence the median angle of the pions to be determined. Although the median angle of the *shower tracks* is observed to depend upon the N_h group, there is no evidence that the median angle of the *pions* shows a similar behaviour. The total multiplicity of the pions for the small N_h group is found to be 2.04 ± 0.36 which, although lower than Fermi's prediction, is in agreement with Daniel *et al.* (1959), who made a more detailed examination of events with $N_h \leq 4$. The uncertainty in the number of protons causes large errors, and a comparison with theory loses much of the significance.

Considering all events, an upper limit to the number of fast protons per star is found to be 0.59 ± 0.06 . For events of small N_h , the corresponding figure seems to be significantly higher, although the uncertainty in the number of protons estimated introduces a relatively greater error.

It is to be noted that the conclusions of the experiment are based on the analysis of shower tracks only, and from events of all N_h grouped together. Slow mesons, of which the number is presumably small, are not considered. Since an upper limit to the proton contamination of the shower tracks could be set up, it was possible to deduce that the meson production occurs as if the incident proton suffered a collision with a

particle of mass no greater than 1.08 ± 0.15 proton mass units. Thus the hypothesis of a nucleon-nucleon collision seems to be justified. Moreover all the fast protons are found to be emitted at angles less than 20° and the majority of them at very small angles.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor C. F. Powell, F.R.S., for the hospitality of his laboratory, and Dr. D. H. Perkins, who suggested this problem and gave valuable advice on many occasions. The author wishes to thank Dr. W. O. Lock of the University of Birmingham for the loan of the plates; the members of the Radiation Laboratory University of California who made the exposure; his colleagues in the laboratory, particularly, Dr. J. W. F. Juritz and Dr. C. J. Waddington for helpful comments; and Dr. R. R. Daniel of the Tata Institute of Fundamental Research, Bombay, for a preprint of a similar investigation, and a personal discussion. He also wishes to thank his wife for some contribution to the scanning effort. A fellowship under the Colombo plan is gratefully acknowledged.

REFERENCES

- APPA RAO, DANIEL, and NEELKANTAN, 1956, *Proc. Indian Acad. Sci. A*, **43**, 181.
- BELIAKOV, VAN SHUFEN, GLAGOLEV, DALKHAZHAY, KIRILLOVA, MARKOV, LEBEDEV, TOLSTOV, TSYGANOV, SHAFRANOVA, JAO TSYNG-SE, BANNIK, BAJAT JAN, GRAMENITSKIJ, DANSYZ, KOSTANASHVILI, LYUBIMOV, NONOFILOV, PODGORETSKI, SKSHIPCHAK, TUVDENDORGE, SHAHULSHVILI, BOGASHEV, BUNYATOV, VISHKI YU, MEREKOV, SIDOROV, 1958, *Cern High Energy Conference Report*, p. 309.
- BISWAS, PETERS, and RAMA, 1955, *Proc. Indian Acad. Sci. A*, **41**, 156.
- CAMERINI, DAVIES, FRANZINETTI, LOCK, PERKINS, and YEKUTIELI, 1951, *Phil. Mag.*, **52**, 1261.
- CAVANAUGH, HASKIN, and SCHIEN, 1955, *Phys. Rev.*, **110**, 1263.
- CESTER, HOANG, and KERNAN, 1956, *Phys. Rev.*, **103**, 1443.
- DANIEL, DAVIES, MULVEY, and PERKINS, 1952, *Phil. Mag.*, **43**, 753.
- DANIEL, KAMESHWAR RAO, MALHOTRA, and TSUZUKI, 1959, Preprint.
- DILWORTH, GOLDSACK, HOANG, and SCARSI, 1953, *Nuov. Cim.*, **10**, 1259.
- DUKE, LOCK, MARCH, MCEVAN, HUGHES, and MUIRHEAD, 1955, *Phil. Mag.*, **46**, 877; 1957, *Ibid.*, **2**, 204.
- EDWARDS, LOSTY, PERKINS, PINKAU, and REYNOLDS, 1958, *Phil. Mag.*, **3**, 237.
- HABER-SCHAIM, 1956, *Nuovo Cim.*, **4**, 669.
- JAPANESE EMULSION GROUP, 1959, *Nuovo Cim. Suppl.*, **11**, 125.
- KALBACH, LORD, and TSAO, 1959, *Phys. Rev.*, **113**, 325.
- KING, 1958, *Phys. Rev.*, **109**, 1344.
- KOBA, 1956, *Proc. VI Rochester Conf. Report*, **4**, 43.
- POWELL, FOWLER, and PERKINS, 1959, *Emulsion Method Atlas* (Pergamon Press), p. 464 (to be published).
- SCHEIN, HASKIN, and GLASER, 1956, *Nuovo Cim.*, **3**, 131.
- SHAPIRO, and STILLER, 1952, *Phys. Rev.*, **87**, 682.

A Statistical Theory of Ionospheric Drifts†

By J. P. DOUGHERTY

Cavendish Laboratory, Cambridge

[Received January 8, 1960]

ABSTRACT

An attempt is made to interpret the horizontal 'drift velocities' observed in ionospheric fading-wave experiments by examining statistically the fluctuations in space and time of the density of ions and electrons. The ionization is assumed to be convected by a turbulent flow in the neutral air, but the magnetic field and the dynamo electric field are also taken into account. The drift velocity to be calculated is defined by means of the space-time correlation function.

For the D region and lower E region, the estimation of this drift velocity is found to be mathematically practicable, using certain assumptions. Results are obtained which show the relative importance of the mean aerodynamic velocity and of the electric field in determining the drift of irregularities in electron density.

For higher regions, the theoretical problem is much less tractable, for reasons discussed here; but it is pointed out that in this case it is unlikely that the irregular behaviour can be ascribed to turbulence anyway, and that the problem of the drift velocity cannot properly be investigated until a satisfactory explanation of the cause of the irregularities is available.

Some remarks are offered on the relation of this theory to the observations and to the dynamo theory.

§ 1. INTRODUCTION

OBSERVATIONS of the fading of radio waves reflected from the ionosphere reveal that the density of ionization has random, or 'irregular' fluctuations superposed on the gradual variation with height. If two or more spaced receivers are used, comparison of the records reveals a noticeable impression of large-scale movement of the irregularities. The procedure for evaluating the velocity from the fading records was given by Briggs *et al.* (1950). Further random movements are superposed on this general 'drift'.

In the E region, since the gas is lightly ionized and collisions are frequent, the movements of the ionization are related to those of the neutral air. It is thought that in this region the air may often be in turbulent motion; if so its movement is described in terms of a mean 'wind' on which are superposed irregular turbulent velocities. It might then be expected (and indeed has sometimes been assumed) that the mean 'wind' in the neutral air is equal to the mean 'drift' of the ionospheric irregularities. This would be the case if the density of the plasma behaved as a simple convected function, like temperature, but unfortunately the Earth's magnetic field

† Communicated by the Author.

makes the ionospheric problem much more complicated than that. The purpose of this paper is to investigate the relation between the two mean velocities. It will be shown that in general they are not the same, and that the relationship between them depends on the height.

Clemmow *et al.* (1955) and Clemmow and Johnson (1959) gave idealized solutions of the relevant equations which showed clearly that results of this kind must be expected, and this investigation began as an attempt to generalize their work. The present approach, however, differs from theirs in several respects. In their formulation, only a uniform motion of the air was permitted; a model 'irregularity' of cylindrical or sinusoidal form was introduced at time $t=0$ (in the plasma only), and its subsequent behaviour in imposed electric and magnetic fields was found. The present paper attempts instead to discuss the problem statistically. The situation is supposed to be one of equilibrium: the turbulent flow continually creates and distorts irregularities, but these, if left to themselves, would gradually be smoothed out by diffusion. It is therefore necessary to make some assumptions about the nature of the air turbulence, and this in itself poses some difficult problems arising from the effect of the buoyancy forces. By restricting attention to the problem of the drift velocity, these assumptions are kept to a minimum, and we shall only consider the simplest which can reasonably be made. This will at least make clear the nature of the problem, and it will also become apparent that to discuss further features of the radio fading, i.e. to discuss the random movements in addition to the statistically defined 'mean drift', would require much deeper analysis. This has recently been initiated by Münch and Wheelon (1958), but so far their discussion of temporal correlations concerns only the components of turbulent velocity and not the density of a plasma convected in a magnetic field.

§ 2. OUTLINE OF THE METHOD

To carry out this programme, we have first to decide upon the assumed form of motion of the neutral air (assuming that the reaction of the plasma on it is negligible). § 3 is taken up with a discussion of this, together with some points concerning the interpretation of the fading pattern on the ground.

The standard dynamical equations for ions and electrons together with Maxwell's equations then determine, in theory, the motion of plasma in external electric and magnetic fields. It is convenient to write all field quantities as a mean part together with a random part whose mean is zero. The mean electric field is then the field of the 'atmospheric dynamo' (Baker and Martyn 1953). This is not genuinely external of course; but it is arbitrary so far as our equations are concerned since we suppose all mean quantities to vary so slowly (with height, with horizontal distance, and with time) that they may be regarded as constant in a sample of space-time large enough to be statistically significant. The random parts are supposed statistically stationary in such a sample, and the equations

governing them are developed in § 5. As in ordinary turbulence theory, it is useful to work with the Fourier transforms of the random quantities, so we derive in § 6 an 'equation of motion' for each Fourier component $N(\mathbf{k}, t)$ of the density fluctuation $n(\mathbf{x}, t)$ of the plasma. This exhibits the temporal changes in $N(\mathbf{k}, t)$ as being due to a 'convection' dependent on the mean wind and electric field, and on the direction of \mathbf{k} ; a 'decay' due to diffusion, also directionally dependent; an 'intermodulation' with other Fourier components via their electrostatic fields; and a complicated process of 'production' by the turbulent flow, which includes amongst other things the mechanism first described by Dungey (1956).

Before entering into these dynamical considerations, however, we give in § 4 the natural definition for the 'drift' velocity of the pattern of a random scalar function in terms of its space-time correlation function. This is a straightforward generalization of that given by Briggs *et al.* (1950) for an isotropic pattern in the amplitude of the signal on the ground. Since the statistical definition of the drift is thereby the same as the one used in analysing the records, it seems reasonable to identify the two, assuming the fading to be due to small structure in the ionosphere, as discussed in § 3.

The results of § 6 are applied in §§ 7 and 8 to the drift so defined; a tentative explicit expression for it for heights up to about 120 km is found assuming an isotropic spectrum, as predicted by Howells (1960). Other literature on turbulent convection (Villars and Weisskopf 1955, Gallet 1955, 1957, Wheelon 1957, Batchelor 1959, and, with the magnetic field included, Howells 1960) concentrates on trying to estimate the spectrum function of $k(=|\mathbf{k}|)$, apparently without general agreement. So it is interesting to note that our result is independent of this. A graphical representation of the way the result depends on height, through the E region, is given, from which one can deduce the relative importance of wind and electric field in determining the drift.

For greater heights, a peculiar difficulty enters and the whole situation becomes rather obscure. The reason for this is discussed in § 9. This also provides an opportunity to make some remarks on the analogous problem for the F region.

A general discussion of the result and its relation with both the radio observations and the dynamo theory is provided in the concluding section.

§ 3. MECHANICS OF THE NON-IONIZED AIR IN THE LOWER IONOSPHERE

As we have already hinted, no adequate theory of the motion of the neutral air in the E region exists—indeed if it were not for the kind of difficulty which it is the purpose of the present paper to try to overcome, one would hope to use the radio observations to help to elucidate some features of the air motion. To make progress we have to adopt some idealized model, so some explanation is due on the question of how likely this is to be near the truth.

Present indications (mainly from data concerning the rate of deformation and diffusion of meteor trails) are that the air motion is to be pictured

somewhat as follows. At the smallest scales, kinetic energy is converted into heat by the action of viscosity and we have a 'dissipation range' in the ordinary way, with buoyancy forces quite unimportant. At rather larger scales, dissipation gradually becomes less important, and buoyancy more important, but there may be an 'equilibrium range' in which neither is very effective, the main feature being the interaction among all pairs of Fourier components of the velocity field, arising from the non-linear inertial term in the Navier-Stokes equation. So far, the energy spectrum function would be roughly isotropic. But at still larger scales the effect of buoyancy becomes the dominant feature, and the spectrum becomes anisotropic. The large 'flat' eddies of about 6 km depth and up to 150 km in horizontal extent, observed by Greenhow and Neufeld (1959) come under this heading. Most of the kinetic energy resides in these eddies. An alternative description of the large-scale motions is obtained by ignoring the non-linear term so that one obtains the linearized theory of 'gravity waves' in a stratified medium (Hines 1959). This gives a useful insight into the general effect of the buoyancy forces and one can easily re-examine the non-linear term later, to establish the error involved, which, if small, can be interpreted as coupling between the various modes. (There is of course another class of motions, associated with compressibility, which at small scales are sound waves, but which at scales comparable with a scale-height are also profoundly modified by buoyancy. This is why Hines has 'two turbulent sequences'.) Finally, on a scale still larger than the large eddies or gravity waves, are the world-wide meteorological motions.

If this description is correct, the usual notion of degradation of energy from larger to smaller scales and finally into the dissipation range, may be used, but care should be exercised in estimating ϵ , the rate of dissipation per unit mass, from the characteristic size L and velocity V of the energy containing eddies. For isotropic turbulence in a uniform fluid one usually writes $\epsilon \approx V^3/L$, but recently reported observations of meteor trails indicate that ϵ is much smaller than V^3/L if the latter refers to the large gravity-dominated eddies. A qualitative explanation appears to be that gravity and the density stratification enable the flow to take on a more ordered structure whereby the excitation of smaller-scale motion through the non-linear terms is less than would be the case for a flow of comparable V and L in a uniform fluid (assuming no special boundary conditions). Thus in our case, to apply the above formula, V and L should refer to the rather smaller-scale (and less energetic) part of the spectrum where buoyancy is just beginning to be dominant, say $V \approx 1$ m/sec, L somewhat less than 1 km, as a very rough estimate. Indeed it might be better not to follow the terminology of the authors just cited, and reserve the word 'turbulence' for the approximately isotropic part of the spectrum, so excluding the gravity waves.

The reader will find more discussion of many of these points in the Proceedings of the International Symposium on Fluid Mechanics in the Ionosphere (*Journal of Geophysical Research*, December 1959) (and the

author is indebted to many of the speakers at this meeting for making clearer these ideas).

This description has been given in some detail partly to emphasize the great complexity of the problem even before introducing the plasma in a magnetic field, and partly to show the significance of a fundamental assumption we shall make. This is to suppose that the radio fading, and the associated measurements of drift velocity, are to be explained in terms of the small scale, roughly isotropic, part of the spectrum. The wavelength for reflection from the E layer is about 100 m, and the size of a Fresnel zone is about 5 km. So any quite small structure in the distribution of plasma density is certainly effective in producing small-scale structure in the radio-wave diffraction pattern on the ground. On the other hand it is quite true that even with large-scale variations only, small-scale structure can appear in the diffraction pattern, as for example in an interference pattern due to specular reflection from a small number of points in a 'wavy' ionosphere. The theoretical problem of interpreting the observed distribution of signal on the ground in terms of its cause in the ionosphere is discussed by Ratcliffe (1956), who points out that the fading pattern does not contain enough information to distinguish between these possibilities. Briggs (private communication) suggests to the author that the highly irregular appearance of records obtained by him (by the rapid frequency-change method) for E region contrast rather markedly with those for the F region, when displayed as a panoramic record. The F region ones are more regular suggesting specular reflection, thus leading one to think that in the E region it is the small-scale fluctuations which are responsible for the fading. If it is subsequently shown that large structure plays an important rôle, it will be necessary to study the convection of plasma by gravity waves in a magnetic field instead of turbulent convection as in this paper. Such a study is in any case relevant to the part of the ionosphere where viscosity is so large that no equilibrium range exists at all, and work on this problem has already been published by Hines (1955, 1956). Estimates of the height at which this might begin to be the case vary; one, due to Stewart, is at about 120 km (remarks at International Symposium, loc. cit.)

Our model, then, is that the flow of neutral air consists of roughly isotropic inertial and dissipation ranges of turbulence, superposed on some uniform mean flow; the latter may incorporate, besides meteorological 'winds', flow due to any gravity waves having a scale larger than a region scanned in making an observation by radio and uniform throughout the time required to make a fading record. Although this is rather an idealization, it enables us to concentrate on the special effect of the magnetic field on the convective process.

§ 4. THE DRIFT VELOCITY OF A RANDOM SCALAR FIELD

Suppose we have an ensemble of realizations of a random scalar $n(\mathbf{x}, t)$, stationary in space and time, with mean zero. Although n will later refer

to the ionization density, it will be convenient here to suppose the variance has been made unity by normalization. Our standard notation for Fourier transforms will be to use the corresponding capital letter, as in

$$n(\mathbf{x}, t) = \int N(\mathbf{k}, t) \exp(i\mathbf{k} \cdot \mathbf{x}) d^3k. \quad . \quad . \quad . \quad . \quad (1)$$

The space-time correlation function is defined by the ensemble average

$$\rho(\boldsymbol{\xi}, \tau) = \overline{n(\mathbf{x}, t)n(\mathbf{x} + \boldsymbol{\xi}, t + \tau)} \quad . \quad . \quad . \quad . \quad (2)$$

and is independent of \mathbf{x} and t (by the stationary property). By Schwartz's inequality this has a maximum (namely unity) at $\boldsymbol{\xi} = 0, \tau = 0$, so assuming a Taylor expansion exists, we have

$$\rho(\boldsymbol{\xi}, \tau) = 1 - \frac{1}{2}[a_{ij}\xi_i\xi_j - 2b_i\xi_i\tau + c\tau^2] + \dots \quad . \quad . \quad . \quad (3)$$

where the quadratic form in brackets must be positive definite. (Summation over i, j , being assumed as usual). The coefficients a_{ij} , b_i , c are given by the second partial derivatives of ρ .

With a suitable ergodic assumption, it is immaterial whether the average in (2) is taken over space, time, or the ensemble of realizations. Imagine, for a particular realization, an observer calculating the 'cross correlation' between the values of $n(\mathbf{x}, t)$ at pairs of points separated by $\boldsymbol{\xi}$ and at two instants t and $t + \tau$, where τ is small; the averaging process being performed only over values of \mathbf{x} . He then selects the value of $\boldsymbol{\xi}$ which *maximizes* the correlation, and defines $\boldsymbol{\xi}/\tau = \mathbf{W}$ to be the drift velocity during the interval. An instantaneous drift is clearly provided by $\lim_{\tau \rightarrow 0} \mathbf{W}$ if the limit exists, which it does if (3) is valid; in fact one easily finds

$$\mathbf{W} = \mathbf{a}^{-1} \cdot \mathbf{b} \quad . \quad . \quad . \quad . \quad (4)$$

where \mathbf{a} is the matrix $[a_{ij}]$ and \mathbf{b} the column-vector $[b_i]$. The correlation for small τ given by (3) when $\boldsymbol{\xi} = \mathbf{W}\tau$ reduces to

$$\rho(\mathbf{W}\tau, \tau) = 1 - \frac{1}{2}[c - \mathbf{b} \cdot \mathbf{a}^{-1} \cdot \mathbf{b}]\tau^2 + \dots \quad . \quad . \quad . \quad (5)$$

so the square bracket is a measure of the random changes which remain after the drift has been removed. For $\boldsymbol{\xi} = \mathbf{W}\tau + \boldsymbol{\eta}$, say, one has

$$\rho(\boldsymbol{\xi}, \tau) = 1 - \frac{1}{2}[a_{ij}\eta_i\eta_j + (c - \mathbf{b} \cdot \mathbf{a}^{-1} \cdot \mathbf{b})\tau^2] + \dots \quad . \quad . \quad . \quad (6)$$

The problem of defining a mean drift was considered by Briggs *et al.* (1950) for the case of random diffraction patterns on the ground produced when radio waves are reflected from the ionosphere. Their approach was conditioned by the fact that one has only a small number of aerials to observe that pattern so ρ can only be obtained for two or three values of $\boldsymbol{\xi}$. If the assumption as to the nature of ρ which they were thereby forced to make is true, their V_x and V_y are the same as would be given by (4) in this two dimensional case. These authors also assumed a spatial isotropy, so $a_{ij} = a\delta_{ij}$, say, and their V_c is defined as $(ac - b^2)^{1/2}/a$ in our notation, with (6) becoming

$$\rho(\boldsymbol{\xi}, \tau) = 1 - \frac{1}{2}a(\boldsymbol{\eta}^2 + V_c^2\tau^2) + \dots \quad . \quad . \quad . \quad (7)$$

When $n(\mathbf{x}, t)$ is given as a Fourier transform, (1), the definition (2) of ρ becomes

$$\begin{aligned}\rho(\xi, \tau) &= \int \int \exp[i\mathbf{k} \cdot (\mathbf{x} + \xi)] \exp(i\mathbf{k}' \cdot \mathbf{x}) \overline{N(\mathbf{k}', t)N(\mathbf{k}, t + \tau)} d^3 k d^3 k' \\ &= \int \int \exp(i\mathbf{k} \cdot \xi) \exp(i\mathbf{k}'' \cdot \mathbf{x}) \overline{N(\mathbf{k}'' - \mathbf{k}, t)N(\mathbf{k}, t + \tau)} d^3 k d^3 k''.\end{aligned}$$

At this point we extend the averaging to include an integration over \mathbf{x} in a large volume $8\pi^3\sigma$, remembering $\rho(\xi, \tau)$ is actually independent of \mathbf{x} , to obtain

$$\begin{aligned}\rho(\xi, \tau) &= (8\pi^3\sigma)^{-1} \int \int \int \exp(i\mathbf{k} \cdot \xi) \exp(i\mathbf{k}'' \cdot \mathbf{x}) \overline{N^*(\mathbf{k} - \mathbf{k}'', t)N(\mathbf{k}, t + \tau)} \\ &\quad \times d^3 k d^3 k'' d^3 x\end{aligned}$$

where we have made use of the reality condition $N(-\mathbf{k}, t) = N^*(\mathbf{k}, t)$. The usual rule for the inversion of Fourier transforms then gives

$$\rho(\xi, \tau) = \sigma^{-1} \int \exp(i\mathbf{k} \cdot \xi) \overline{N^*(\mathbf{k}, t)N(\mathbf{k}, t + \tau)} d^3 k. \quad (8)$$

Comparing with (3) and (4), we have

$$\left. \begin{aligned}a_{ij} &= -\left(\frac{\partial^2 \rho}{\partial \xi_i \partial \xi_j}\right)_0 = \sigma^{-1} \int k_i k_j \overline{|N(\mathbf{k}, t)|^2} d^3 k, \\ b_i &= \left(\frac{\partial^2 \rho}{\partial \xi_i \partial \tau}\right)_0 = i\sigma^{-1} \int k_i \overline{N^*(\mathbf{k}, t) \partial N(\mathbf{k}, t) / \partial t} d^3 k.\end{aligned}\right\} \quad (9)$$

These are of course actually independent of t , since $\overline{|N(\mathbf{k}, t)|^2}$ is so; this also shows incidentally that \mathbf{b} as given by (9) is real, as it should be.

The above work uses Fourier integrals formally, without regard for convergence. Rigorous exposition of manipulations of this type can be given, provided the Stieltjes integral notation is used, see for instance Moyal (1952). We shall recover finite expressions on evaluating $\mathbf{a}^{-1} \cdot \mathbf{b}$.

§ 5. DYNAMICS OF THE IONS AND ELECTRONS

Detailed considerations of the effects likely to be important in the type of disturbance to be investigated, have been given by Dungey (1956) and by Clemmow and Johnson (1959) so need not be repeated here. We shall assume:

(i) Fluctuations in the magnetic field are to be neglected. A convenient choice of axes makes $\mathbf{B} = (0, 0, B)$.

(ii) The plasma consists of electrons, and ions of one kind only, singly charged, and the dynamical processes take place so rapidly that ionization and recombination processes can be neglected. Inertial and gravitational terms are also neglected.

(iii) The number densities of ions and electrons remain almost exactly equal, the difference being much smaller even than the *perturbations* in number density. (Clemmow and Johnson do not assume this at the outset, but the result of their linearized calculation is that additional modes of

oscillation exhibiting non-neutrality, but decaying very rapidly, enter the theory; they are neglected here.) This is the usual procedure, (see Spitzer 1956, p. 16) which is valid provided the length-scales of the problem are all much greater than the Debye distance, which for the ionosphere is a few millimetres. The number of variables is thus reduced by one, but Poisson's equation need not be used. This is, however, contrary to Booker's (1956) treatment, which introduces both 'ionic' and 'electronic' irregularities.

We shall write script letters for the following mean quantities: \mathcal{N} for the density of ions and electrons, $\mathcal{V}^+, \mathcal{V}^-$ for their velocities, and \mathcal{U} for the velocity of the neutral air. The random parts are denoted by the corresponding small letters $n, \mathbf{v}^+, \mathbf{v}^-, \mathbf{u}$. But the mean electric field is written \mathbf{E} , and the random part $-\nabla\phi$. Capital letters N, Φ , are the Fourier transforms as in (1). We shall frequently write symbols or whole equations with alternative signs \pm , for brevity, to indicate ions and electrons respectively. Thus $m^\pm, \pm e$ are the mass and charge, while ν^\pm are the frequencies of collisions with neutral particles (ion-electron collisions being neglected). We assume partial pressures $\mathcal{N}kT$ for ions and electrons at constant temperature T . The gyro-frequencies are $\omega^\pm = eB/m^\pm$, and as in Baker and Martyn (1953) the useful angles α are defined by

$$\tan \alpha^\pm = \omega^\pm / \nu^\pm. \quad (10)$$

The general equations of continuity and momentum are

$$\frac{\partial \mathcal{N}}{\partial t} = -\operatorname{div}(\mathcal{N}\mathcal{V}^\pm), \quad (11)$$

$$m^\pm \nu^\pm (\mathcal{V}^\pm - \mathcal{U}) \mp e \mathcal{V}^\pm \times \mathbf{B} = \pm e \mathbf{E} - \frac{kT}{\mathcal{N}} \nabla \mathcal{N}. \quad . . . (12)$$

which together with $\operatorname{curl} \mathbf{E} = 0$ determine the motion of the plasma. Equation (12) may be solved for \mathcal{V}^\pm in the form

$$\mathcal{V}^\pm = \mathbf{M}^\pm \left(\frac{\pm e}{m^\pm \nu^\pm} \mathbf{E} + \mathcal{U} - \frac{kT}{m^\pm \nu^\pm \mathcal{N}} \nabla \mathcal{N} \right) \quad (13)$$

where the matrices \mathbf{M}^\pm are defined by

$$\mathbf{M}^\pm = \begin{bmatrix} \cos^2 \alpha^\pm & \pm \sin \alpha^\pm \cos \alpha^\pm & 0 \\ \mp \sin \alpha^\pm \cos \alpha^\pm & \cos^2 \alpha^\pm & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad . . . (14)$$

Since the script letters and \mathbf{E} were defined as the steady mean values, all the derivatives in (11)–(13) are actually zero; we retained them formally so that the equations in the disturbed state are then obtained by replacing \mathcal{U} by $\mathcal{U} + \mathbf{u}$, etc., and \mathbf{E} by $\mathbf{E} - \nabla\phi$. One then subtracts the equations satisfied by the mean quantities,

$$\mathcal{V}^\pm = \mathbf{M}^\pm \left(\frac{\pm e}{m^\pm \nu^\pm} \mathbf{E} + \mathcal{U} \right) \quad (15)$$

to obtain

$$\mathbf{v}^\pm = \mathbf{M}^\pm \left(\mp \frac{e}{m^\pm \nu^\pm} \nabla\phi + \mathbf{u} - \frac{kT}{m^\pm \nu^\pm (\mathcal{N} + n)} \nabla n \right) \quad . . . (16)$$

so using (11), and going into suffix notation,

$$\begin{aligned}
 \frac{\partial n}{\partial t} &= -\frac{\partial}{\partial x_i} [n\mathcal{V}_i^\pm + (\mathcal{N} + n)v_i^\pm] \\
 &= -\mathcal{V}_i^\pm \frac{\partial n}{\partial x_i} - M_{ij}^\pm \frac{\partial}{\partial x_i} \left\{ \mp \frac{e(\mathcal{N} + n)}{m^{\pm\nu\pm}} \frac{\partial \phi}{\partial x_j} - \frac{kT}{m^{\pm\nu\pm}} \frac{\partial n}{\partial x_j} + (\mathcal{N} + n)u_j \right\} \\
 &= -\mathcal{V}_i^\pm \frac{\partial n}{\partial x_i} + M_{ij}^\pm \left\{ \pm \frac{e(\mathcal{N} + n)}{m^{\pm\nu\pm}} \frac{\partial^2 \phi}{\partial x_i \partial x_j} \pm \frac{e}{m^{\pm\nu\pm}} \frac{\partial n}{\partial x_i} \frac{\partial \phi}{\partial x_j} \right. \\
 &\quad \left. + \frac{kT}{m^{\pm\nu\pm}} \frac{\partial^2 n}{\partial x_i \partial x_j} - u_j \frac{\partial n}{\partial x_i} - (\mathcal{N} + n) \frac{\partial u_j}{\partial x_i} \right\}. \quad (17)
 \end{aligned}$$

Except in the non-magnetic case, when \mathbf{M}^\pm become unit matrices, it is impossible to eliminate explicitly the internal potential ϕ which couples the two simultaneous diffusion equations (17) and makes the problem non-linear. We are, however, only concerned to find enough about the solutions of (17) to make an estimate of the drift velocity \mathbf{W} of § 4, and to do this the equation will be converted into one satisfied by the Fourier transforms.

§ 6. THE EQUATION OF MOTION OF A FOURIER COMPONENT

One simple approximation in eqns. (17) will be made at once: the fractional changes in density are known observationally to be quite small, though the density *gradients* involved are larger than the steady gradient with height which we have already neglected. Hence we can ignore n in comparison with \mathcal{N} in factors where n is not differentiated, i.e. we replace the two brackets $(\mathcal{N} + n)$ by \mathcal{N} in (17).

After doing this and substituting expressions of type (1) into (17), we obtain, using some straightforward changes of variable, two coupled integral equations for $N(\mathbf{k}, t)$ and $\Phi(\mathbf{k}, t)$:

$$\begin{aligned}
 \frac{\partial N(\mathbf{k}, t)}{\partial t} &= -i\mathcal{V}^\pm \cdot \mathbf{k} N(\mathbf{k}, t) + M_{ij}^\pm \left\{ \mp \frac{\mathcal{N}e}{m^{\pm\nu\pm}} k_i k_j \Phi(\mathbf{k}, t) - \frac{kT}{m^{\pm\nu\pm}} k_i k_j N(\mathbf{k}, t) \right. \\
 &\quad \left. \mp \frac{e}{m^{\pm\nu\pm}} \int k'_i (k_j - k'_j) N(\mathbf{k}', t) \Phi(\mathbf{k} - \mathbf{k}', t) d^3 k' \right. \\
 &\quad \left. - i \int k'_i U_j(\mathbf{k} - \mathbf{k}', t) N(\mathbf{k}', t) d^3 k' - i\mathcal{N} k_i U_j(\mathbf{k}, t) \right\}. \quad (18)
 \end{aligned}$$

The terms on the right-hand side of (18) have the following interpretation. The first represents a convection of the Fourier components of wave number \mathbf{k} in the ions or electrons with the mean drifts \mathcal{V}^+ or \mathcal{V}^- respectively; the first in the large bracket is the effect of the Fourier component with the same \mathbf{k} in the electrostatic field, set up to prevent charge separation. The term involving kT gives the effect of molecular diffusion in smoothing out the Fourier component \mathbf{k} , while the last term of all can be identified as Dungey's mechanism (Dungey 1956) for producing irregularities in the plasma by means of a magnetic field. We note that in the absence of the field, M_{ij}^\pm reduce to δ_{ij} , and the term

becomes small because $\mathbf{k} \cdot \mathbf{U}(\mathbf{k}, t) \approx 0$ since the turbulent velocities are much smaller than the sound velocity; the term then represents the compressions in the neutral air being repeated in the plasma, an effect rejected by Villars and Weisskopf (1955) and by Gallet (1955) as being too small to explain the strength of irregularity observed. (Maeda *et al.* 1959, suggest similarly that the compressions are repeated in the plasma via the ionizing process rather than mechanically: this seems to the present writer to be far too small an effect, for similar reasons).

To complete our description of the right-hand side of (18), there are the two integral terms representing intermodulation between all pairs of Fourier components \mathbf{k} and \mathbf{k}' . In the first case it is through the internal electric field, while the second integral is a generalization of that arising in the non-magnetic case, whereby the turbulent velocity field acting on the local gradients in $n(\mathbf{x}, t)$ due to the existing irregularities, continually redistributes irregularity throughout the spectrum $N(\mathbf{k}, t)$.

The gentle gradient in the mean N is also involved in this last mentioned process, but (as emphasized by Batchelor (1959, p. 118) but not appreciated by some of the earlier writers) the gradients due to the irregularities themselves are much more important, except at the smallest wave numbers. So (18) is still valid in the dissipation and equilibrium ranges, but omits a term describing the 'input' at small wave number, which however is immaterial to us in what follows.

It will be seen from (18) that the difficulty of eliminating ϕ between the ion- and electron-equations is at least as great as before; we have removed spatial differentiations at the price of introducing integral equations. Nevertheless (18) supplies the equation which is sufficient for the present purpose, namely that obtained by eliminating the term in which $\Phi(\mathbf{k}, t)$ appears explicitly, without regard for the term in which $\Phi(\mathbf{k} - \mathbf{k}', t)$ appears under integration. If θ is the angle between \mathbf{k} and \mathbf{B} , so that

$$\cos \theta = k_3/k \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

and introducing

$$S^\pm = M_{ij}^\pm k_i k_j / k^2 = \cos^2 \alpha^\pm \sin^2 \theta + \cos^2 \theta \quad . \quad . \quad . \quad . \quad (20)$$

the appropriate linear combination of equations (18) is

$$\begin{aligned} \frac{\partial N(\mathbf{k}, t)}{\partial t} = & -i \mathbf{W}(\theta) \cdot \mathbf{k} N(\mathbf{k}, t) - D \chi(\theta) k^2 N(\mathbf{k}, t) \\ & + P_{ij}(\theta) \int k_i' (k_j - k_j') N(\mathbf{k}', t) \Phi(\mathbf{k} - \mathbf{k}', t) d^3 k' \\ & - Q_{ij}(\theta) \left\{ i \int k_i' U_j(\mathbf{k} - \mathbf{k}', t) N(\mathbf{k}', t) d^3 k' + i \mathcal{A} k_i U_j(\mathbf{k}, t) \right\} . \end{aligned} \quad . \quad . \quad . \quad (21)$$

The following abbreviations are used in (21):

$$D = \frac{2kT}{m^+ v^+ + m^- v^-} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

This is the standard "ambipolar" diffusion coefficient, and

$$\mathbf{W}(\theta) = \frac{S^+ \tan \alpha^+ \mathcal{V}^- + S^- \tan \alpha^- \mathcal{V}^+}{S^+ \tan \alpha^+ + S^- \tan \alpha^-}, \quad . \quad . \quad . \quad (23)$$

$$\chi(\theta) = \frac{S^+ S^- (\tan \alpha^+ + \tan \alpha^-)}{S^+ \tan \alpha^+ + S^- \tan \alpha^-}, \quad . \quad . \quad . \quad (24)$$

$$\mathbf{P}(\theta) = \frac{e^2 B (S^+ \mathbf{M}^- - S^- \mathbf{M}^+)}{m^+ \nu^+ m^- \nu^- (S^+ \tan \alpha^+ + S^- \tan \alpha^-)}, \quad . \quad . \quad . \quad (25)$$

$$\mathbf{Q}(\theta) = \frac{S^+ \tan \alpha^+ \mathbf{M}^- + S^- \tan \alpha^- \mathbf{M}^+}{S^+ \tan \alpha^+ + S^- \tan \alpha^-}. \quad . \quad . \quad (26)$$

Equation (21) can be regarded as differing from (18) in that a common velocity of convection, $\mathbf{W}(\theta)$ has been achieved to replace \mathcal{V}^\pm , and a common diffusive decay as if due to a diffusivity $D\chi(\theta)$. The terms with \mathbf{P} and \mathbf{Q} can still be regarded as describing Dungey's process, together with the intermodulation of different Fourier components, but in practice, lack of knowledge of $\Phi(\mathbf{k}, t)$ would prevent any exact use of (21) alone.

$\mathbf{W}(\theta)$ is clearly identical with the 'wave velocity' found by Clemmow and Johnson (1959) for the linearized theory in the absence of turbulence. It can be regarded as a 'compromise' drift, between \mathcal{V}^+ and \mathcal{V}^- ; being in fact a weighted mean of them, the weight depending on θ and on α^\pm . It should not be confused with the \mathbf{W} (written without an argument) defined in § 4, though as we shall see, they are closely related; if it were possible for all the contributions to $N(\mathbf{k}, t)$ to arise only from those values of \mathbf{k} corresponding to a fixed value of θ , \mathbf{W} would then be just $\mathbf{W}(\theta)$ evaluated for that θ . Note also that $\chi(\theta)$ is unity for $\theta = 0$, and decreases monotonically to its value at $\theta = \frac{1}{2}\pi$.

§ 7. INTEGRAL EXPRESSIONS FOR THE DRIFT VELOCITY

Although we have now set up a complete formal scheme for tackling the problem, namely eqns. (4), (9) and (18), it is already clear that even if we had enough information about the behaviour of $\mathbf{U}(\mathbf{k}, t)$, an exact use of the equations would be prohibitively complex. But besides this, little is known about the temporal behaviour of the velocity field in a turbulent flow, even in the absence of buoyancy forces. Because of the non-linearity of the Navier-Stokes equation, the equation satisfied by the $\mathbf{U}(\mathbf{k}, t)$ for each realization is of a similar intractable nature, and has not been much used (but see Münch and Wheelon 1958). Clearly, then, we shall have to make some approximations or statistical assumptions if we are to make any estimate of the overall drift \mathbf{W} . For this reason our own estimate of it must be regarded as tentative.

The integrals (9) which we have to evaluate, extend over the whole of \mathbf{k} -space. Although we shall continue to write them conventionally in this way, it should be noted that our assumption discussed in § 3 makes the contribution from the large scales (small values of \mathbf{k}) negligible. This

is now seen to be even more plausible on account of the factors k^2 and $k\partial/\partial t$ in the integrands.

As pointed out by Wheelon (1957) the diffusivity D is about twice the kinematic viscosity in this context. Hence the convective equilibrium range terminates roughly where the inertial one does, and the cutoff beyond is almost certainly very rapid (Batchelor 1959). So the dissipation range may also be omitted from our integrations as a first approximation.

By retaining the range of integration as the whole of \mathbf{k} -space we are really supposing that the temporal behaviour of all Fourier components is of the same form as for those in the equilibrium range, on the grounds that those parts of the range for which this is not so are unimportant anyway.

Consider now the first of the integrals (9). To evaluate this we need only the 'spectrum density' $|\overline{N(\mathbf{k}, t)}|^2$. This has been considered recently by Howells (1960), who concludes that despite the apparent anisotropy of the situation, the spectrum is roughly isotropic in the equilibrium range. Two cases arise: (i) below about 125 km height, where $\nu^+ > \omega^+$, i.e. $\tan \alpha^+ < 1$; (ii) above that level, where the opposite holds. Case (i) has been worked out in more detail, and one can say with confidence that the spectrum is close to isotropic. Even for case (ii) the departure from isotropy appears to be fairly small. (The anisotropy of the diffusion is irrelevant here because we ignore the range in which any diffusion is important. The highly elongated irregularities suggested by Booker (1956) are therefore unlikely to occur in the parts of the ionosphere where there is local turbulence, though they may occur where irregularities exist for some other reason).

We can, then, write $|\overline{N(\mathbf{k}, t)}|^2 = g(k)$, say. Using polar angles θ, ψ , relative to the 03 axis, (9) gives (where $\mathbf{l} = \mathbf{k}/k$)

$$a_{ij} = \sigma^{-1} \int_0^\infty k^4 g(k) dk \int_{\theta=0}^\pi \int_{\psi=0}^{2\pi} l_i l_j \sin \theta d\theta d\psi, \\ = G \delta_{ij}, \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where

$$G = \frac{4\pi}{3\sigma} \int_0^\infty k^4 g(k) dk. \quad . \quad . \quad . \quad . \quad . \quad (28)$$

Turning to the second of the integrals (9), we must use the work of §6 to estimate $\partial N(\mathbf{k}, t)/\partial t$, and we shall use the form (21) in preference to the simultaneous eqns. (18). It is helpful to picture $N(\mathbf{k}, t)$, for each fixed \mathbf{k} , as a locus in the complex plane. Then the first term of (21) is a tendency for the 'vector' representing N to rotate at a definite angular velocity $-\mathbf{k} \cdot \mathbf{W}(\theta)$; the second term need not be considered. The two integral terms are of a stochastic nature and represent some kind of 'random walk' process superposed on the rotation. The last term, if it is important, is of a more systematic nature than the integrals, because at fairly small scales $U_j(\mathbf{k}, t)$ varies nearly systematically—the usual concept of convection of the small eddies by the mean or large-scale flow makes U tend to rotate with angular frequency $-\mathbf{k} \cdot \mathbf{U}$.

Using (18) to write eqn. (9) free of time derivatives, we obtain

$$\mathbf{b} = \sigma^{-1} \int \mathbf{k} [\mathbf{k} \cdot \mathbf{W}(\theta)] \overline{|N(\mathbf{k}, t)|^2} d^3 k, \quad (29)$$

$$- i D \sigma^{-1} \int \chi(\theta) \mathbf{k} k^2 \overline{|N(\mathbf{k}, t)|^2} d^3 k, \quad (30)$$

$$\left. \begin{aligned} & + i \sigma^{-1} \int \int \mathbf{k} k_i' (k_j - k_j') P_{ij}(\theta) \overline{N^*(\mathbf{k}, t) N(\mathbf{k}', t) \Phi(\mathbf{k} - \mathbf{k}', t)} d^3 k d^3 k' \\ & + \sigma^{-1} \int \int \mathbf{k} k_i' Q_{ij}(\theta) \overline{N^*(\mathbf{k}, t) N(\mathbf{k}', t) U_j(\mathbf{k} - \mathbf{k}', t)} d^3 k d^3 k', \end{aligned} \right\} \quad (31)$$

$$+ \sigma^{-1} \mathcal{N} \int \mathbf{k} k_i Q_{ij}(\theta) \overline{N^*(\mathbf{k}, t) U_j(\mathbf{k}, t)} d^3 k. \quad (32)$$

Here θ refers to the polar angle for the wave number \mathbf{k} , not \mathbf{k}' , and it is convenient to number separately parts of the right-hand side.

For the part (29), continuing our assumption of isotropy, integration over $|\mathbf{k}|$ in polar coordinates and making use of the definition (28) of G yields a contribution to $\mathbf{W} (= \mathbf{a}^{-1} \cdot \mathbf{b})$ which is

$$\frac{3}{4\pi} \int_{\psi=0}^{2\pi} \int_{\theta=0}^{\pi} \mathbf{l} [\mathbf{l} \cdot \mathbf{W}(\theta)] \sin \theta d\theta d\psi. \quad (33)$$

The part (30) is zero since the integrand is an odd function of \mathbf{k} .

The part (31) shows explicitly the contribution to drift velocity of the intermodulation effects. Triple correlations of N , Φ and \mathbf{U} evaluated at three different wave numbers are involved here. We shall assume that these correlations are negligible; or in other words that the interactions between different wave numbers are entirely random, showing no preferred direction in space-time. This is not to say that the non-linear effects are unimportant—they are indeed important in determining the form of the spectrum, i.e. $g(k)$, but can be ignored in the drift problem.

Finally, we have to examine part (32). For case (i) of the classification mentioned above, i.e. for heights rather less than 125 km, this term does provide a systematic correlation, but it is small, because Dungey's effect (see Dungey 1956) is itself small in this region. This can be seen explicitly in our formulation by means of eqn. (26), which shows \mathbf{Q} as a weighted mean of \mathbf{M}^+ and \mathbf{M}^- . Below about 90 km, where $\nu > \omega$ for the electrons also, both \mathbf{M}^\pm are very nearly the unit matrix. Between 90 and 125 km, where \mathbf{M}^+ is still nearly unit, \mathbf{M}^- is not so, however the factor $\tan \alpha^-$ increases to such an extent that except for θ very close to $\frac{1}{2}\pi$, the mean is weighted heavily in favour of \mathbf{M}^+ . Thus it is only near and above 125 km that \mathbf{Q} differs appreciably from the unit matrix, and until that happens (32) gives zero since $\mathbf{k} \cdot \mathbf{U}(\mathbf{k}, t) \approx 0$. Since $\mathbf{U}(\mathbf{k}, t)$ is isotropic, no special weight is attached to the small range of θ in which this argument fails, so as a first approximation, we can justifiably neglect (32) for case (i).

For case (ii), the position with regard to (32) is much more difficult, but some remarks are offered in § 9.

§ 8. EVALUATION OF THE DRIFT VELOCITY IN THE SIMPLE CASE

For case (i), the only surviving contribution to \mathbf{W} , according to the foregoing arguments, is given by (33). This is no more than a mean over all directions of the velocities for the Fourier components, each resolved along the 'wave normal'. Our statistical definition of drift has thus given a new significance to Clemmow and Johnson's 'wave velocity', originally derived by a linearized theory in the absence of turbulence. Note also that the distribution among wave numbers, $g(k)$, disappears from the drift problem once an isotropic spectrum has been assumed. This is of course to be traced back to the fact that the convection for any Fourier component \mathbf{k} depends solely on θ and not on the wave number k . Since there has been some controversy (summarized by Batchelor 1959) concerning $g(k)$ even for a non-magnetic case, we can regard this as being rather fortunate.

The last step is to write $\mathbf{l} = (\sin \theta \cos \psi, \sin \theta \sin \psi, \cos \theta)$ and use (20) and (23) to carry out the integrations (33); this involves only elementary integrations, and a convenient form of the result is

$$W_{1,2} = (1 - R_1) \mathcal{V}_{1,2}^+ + R_1 \mathcal{V}_{1,2}^- \quad \text{and} \quad W_3 = (1 - R_3) \mathcal{V}_3^+ + R_3 \mathcal{V}_3^-. \quad (34)$$

Thus each component of \mathbf{W} is a weighted mean of the corresponding components of \mathcal{V}^\pm , the weighting factors being R_1 for those perpendicular to the field, and R_3 for that parallel to it. R_1 and R_3 depend solely upon α^\pm , i.e. upon height in the ionosphere, according to the formulae

$$R_j = \frac{\sin \alpha^+ \cos \alpha^-}{\sin (\alpha^+ + \alpha^-)} \{1 + \tan \alpha^- \tan (\alpha^- - \alpha^+) I_j [\cos \alpha^+ \cos \alpha^- \cos (\alpha^- - \alpha^+)]\} \quad \dots \quad (35)$$

(for $j = 1, 3$)

$$\begin{aligned} \text{where} \quad I_1(x) &= \frac{3x}{2} \int_0^{\pi/2} \frac{\sin^5 \theta d\theta}{\cos^2 \theta + x \sin^2 \theta} \\ &= \frac{x}{2(1-x)^2} \left\{ 2x - 5 + \frac{3}{x^{1/2}(1-x)^{1/2}} \tan^{-1} \left[\frac{1-x}{x} \right]^{1/2} \right\}. \end{aligned} \quad (36)$$

$$\text{and} \quad I_3(x) = 2x[1 - I_1(x)]. \quad \dots \quad (37)$$

For small x one uses

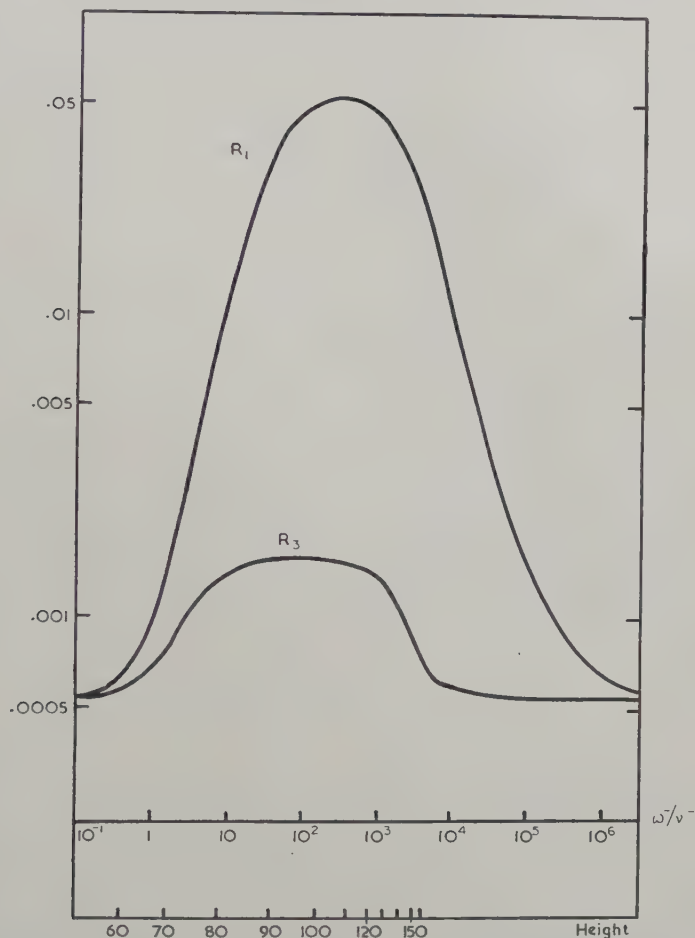
$$I_1(x) \approx \frac{3\pi}{4} x^{1/2} - 4x + \frac{9\pi}{16} x^{3/2} + \dots$$

The first factor of (35) can also be written

$$\frac{\sin \alpha^+ \cos \alpha^-}{\sin (\alpha^+ + \alpha^-)} = [1 + \tan \alpha^- / \tan \alpha^+]^{-1} = [1 + m^+ \nu^+ / m^- \nu^-]^{-1} \approx m^- \nu^- / m^+ \nu^+, \quad \dots \quad (38)$$

so is approximately independent of height in the ionosphere since ν^+ / ν^- is nearly constant—we shall take it to be $1/30$ in the numerical work which follows, and m^+ / m^- to be 5.5×10^4 . Then R_1 and R_3 depend only on say

$\tan \alpha^- = \omega^-/\nu^-$, since α^+ is then fixed. The figure shows the values calculated from (35). This can be applied for any atmospheric model if the two numerical parameters are as just quoted, but estimated heights are also marked on the horizontal axis, taking $\omega^- = 7.5 \times 10^6 \text{ sec}^{-1}$ and values of ν^- based on recent atmospheric models (kindly supplied by Dr. K. Weekes).



Illustrating the dependence upon height of the weighting factors R_1 and R_3 which give the relative importance of the ion-drift and electron-drift in determining the drift of the irregularities.

Consider first the lowest levels, around 70 km, where both α^\pm are small. The large bracket factor in (35) is then ≈ 1 , (for both $j = 1, 3$) and we have $R_1 \approx R_3 \approx [1 + m^+\nu^+/m^-\nu^-]^{-1}$. But (15) has unit matrices in place of \mathbf{M}^\pm in this approximation, so that, as expected, equations (33) revert to $\mathbf{W} = \mathbf{U}$, irrespective of \mathbf{E} .

Slightly higher in the ionosphere this simple result changes, partly because (15) becomes more complicated and partly because the coefficients R_1 and R_3 increase, especially R_1 . One might have expected intuitively

that however \mathcal{V}^\pm vary in this way, the drift of irregularities would always remain approximately equal to \mathcal{V}^+ on account of the much greater mass of the ions. The coefficients R_j measure just the extent to which this must be corrected due to the motion of the electrons, an effect reaching a maximum in the middle of the region characterized by $\omega^- > \nu^-$ but $\omega^+ < \nu^+$, i.e. precisely the dynamo region. The effect is, however, fairly small unless \mathbf{E} and \mathcal{U} are so arranged that \mathcal{V}^- is somewhat larger than \mathcal{V}^+ . It can also be interpreted as the dynamical effect of the 'polarization' of the irregularities due to the mean current flow and the fluctuations in conductivity (cf. Clemmow and Johnson's example of a cylinder, 1955).

Above the dynamo region, the coefficients R_j decrease again to their former value, but the further complications discussed in the next section make this simple theory unreliable.

§ 9. THE UPPER E REGION AND THE F REGION

Above the dynamo region, \mathcal{V}^\pm tend to equality, in which case the velocity given by (33) is also the same, namely

$$\mathbf{W} \approx \mathcal{V}^+ \approx \mathcal{V}^- \approx (E_2/B, -E_1/B, \mathcal{U}_3) \quad . \quad . \quad . \quad (39)$$

the coefficients R_j being practically irrelevant. However, as \mathbf{Q} is no longer nearly the unit matrix, the contribution from part (32) of the expression for \mathbf{b} must be considered, and it has not proved possible to derive any mathematical results.

Some insight into this situation may be gained by linearizing (21), and omitting the diffusion term, for simplicity. We then have a 'small signal' theory, in which each Fourier component behaves independently and Dungey's effect may be studied in isolation. We have, for each \mathbf{k} ,

$$\frac{\partial N}{\partial t} + i \mathbf{W}(\theta) \cdot \mathbf{k} N = -i \mathcal{N} Q_{ij} k_i U_j, \quad . \quad . \quad . \quad (40)$$

Taking U_j to be proportional to $\exp(-ipt)$ where $p = \mathbf{k} \cdot \mathcal{U}$, which as mentioned above is a good approximation for the smaller eddies, the solution for $N(t)$ consists of the particular integral proportional to $\exp(-ipt)$ and a complementary function proportional to $\exp[-i\mathbf{k} \cdot \mathbf{W}(\theta)t]$. The former gives a finite correlation in both (29) and (32), and the latter as before in (29). But the amplitude of the complementary function is arbitrary in a linearized theory; its mean equilibrium amplitude in any situation to which the linear theory may approximate is determined by the non-linear terms. Setting all the complimentary functions zero leads of course to $\mathbf{W} = \mathcal{U}$. Presumably when both turbulent mixing and Dungey's process are important, \mathbf{W} is in some sense intermediate between \mathcal{U} and the value found in § 8.

As the height of the maximum electron density in the E region generally occurs at 120–130 km, no reflections can be observed from this level up to something like 180 km, and it seems most unlikely that turbulence occurs at or above this latter height. So in practice the difficulty discussed in this section is probably rarely serious.

The main point of interest in this discussion is that it has emerged that there is no unique 'velocity of irregularities' at a given level in the ionosphere. It is necessary to enquire into the details of the process producing the irregularity—in the present case we are unable to proceed without knowing in greater detail the rôles of turbulent mixing and Dungey's process. Similarly our theory for the lower E region assumes turbulent mixing as the cause, but the whole question would need to be reconsidered if, for example it were proposed that inhomogeneities in the ionizing radiation produced irregular electron density.

In the case of the F region, the difficulty becomes acute, as there is no adequate theory of the origin of the observed irregularities. Dagg (1957 a) has reviewed several explanations and criticized all of them. His own theory (1957 b), of small-scale electric fields originating in the dynamo region, is also open to criticism, however, as such electric fields would not produce compressions in the plasma (Dougherty 1959). Alfvén waves travelling along lines of force and originating in the outer ionosphere have also been suggested, (Dessler 1958) but the velocity would then be the wave velocity, which is too large to agree with the observations.

§ 10. DISCUSSION OF RESULTS

Let us now reconsider our objective, the interpretation of the drift velocities derived from fading records. To do this, we work in new axes $0x$, $0y$ and $0z$, taken along magnetic East, North, and vertically upwards, respectively. Then assuming that the fading is imposed at or near the height of reflection, the spaced-receiver experiments yield W_x and W_y appropriate to that height.

Below about 90 kms, we can with confidence say that this gives directly the wind velocity for the neutral air, \mathcal{U}_x and \mathcal{U}_y . (\mathcal{U}_z can generally be assumed to be an order of magnitude smaller.) This is true even if a magnetic field is present.

In the neighbourhood of the dynamo layer, the electric field plays an important part in determining \mathbf{W} , and one cannot proceed without ambiguity from W_x, W_y to $\mathcal{U}_x, \mathcal{U}_y$. Taking $\mathcal{U}_z = 0$ and assuming E_z adjusted to make $\mathcal{V}_z^+ = \mathcal{V}_z^-$ (no vertical current), we still have two further unknown quantities, E_x, E_y .

The further relations between $\mathcal{U}_x, \mathcal{U}_y$ and E_x, E_y must of course be provided by means of a dynamo theory of their variation over the whole globe. So the present theory can be used to find whether a proposed dynamo theory successfully predicts the drifts, and for example, their diurnal behaviour.

It is certainly possible to construct numerical examples in which the error involved in identifying the components of \mathbf{W} as those of \mathcal{U} would be very serious.

For much of the published data (see for example Briggs and Spencer 1954) the height to which the drift refers was not measured, except so far as to identify the region, E or F, and this makes satisfactory interpretation virtually impossible. Recently, however, Jones (1958) has reported

results in which the height was estimated; simultaneous measurements at different heights near the dynamo region revealed a gradient in the velocity, which Jones expressed as a change of phase in the diurnal component of drift. A possible explanation is that although \mathcal{U} and \mathbf{E} are each independent of height (as assumed in the dynamo theory) the diurnal variations of \mathcal{U}_x , \mathcal{U}_y , E_x , E_y , have different phases. Since \mathbf{W} depends increasingly upon \mathbf{E} with increasing height, a gradual change in phase of \mathbf{W} would automatically take place.

In conclusion, we may claim that a theoretical basis has been laid for relating the ionospheric drift observations to the winds and electric fields of the dynamo region.

ACKNOWLEDGMENTS

I have enjoyed many stimulating discussions with Mr. J. A. Ratcliffe, Dr. P. C. Clemmow, Dr. K. Weekes, Dr. K. G. Budden, Dr. B. H. Briggs and Dr. I. D. Howells, and my thanks are due to them all.

This work has been carried out during the tenure of a D.S.I.R. Research Studentship.

My attendance at the International Symposium on Fluid Mechanics in the Ionosphere, at Cornell University, was made possible by the United States Office of Naval Research.

REFERENCES

- BAKER, W. G., and MARTYN, D. F., 1953, *Phil. Trans. A*, **246**, 281.
 BATCHELOR, G. K., 1959, *J. Fluid Mech.*, **5**, 113.
 BOOKER, H. G., 1956, *J. geophys. Res.*, **61**, 673.
 BRIGGS, B. H., PHILLIPS, G. J., and SHINN, D. H., 1950, *Proc. phys. Soc. Lond. B*, **63**, 106.
 BRIGGS, B. H., and SPENCER, M., 1954, *Rep. Progr. Phys.*, **17**, 245.
 CLEMMOW, P. C., and JOHNSON, M. A., 1959, *J. atmos. terr. Phys.*, **16**, 21.
 CLEMMOW, P. C., JOHNSON, M. A., and WEEKES, K., 1955, *Rep. Ionosph. Conf.* (London: Physical Society), p. 136.
 DAGG, M., 1957 a, *J. atmos. terr. Phys.*, **11**, 133; 1957 b, *Ibid.*, **11**, 139.
 DESSLER, A. J., 1958, *J. geophys. Res.*, **63**, 507.
 DOUGHERTY, J. P., 1959, *J. geophys. Res.*, **64**, 2215.
 DUNGEY, J. W., 1956, *J. atmos. terr. Phys.*, **8**, 39.
 GALLET, R. M., 1955, *Proc. Inst. Radio Engrs. N.Y.*, **43**, 1240; 1957, *Polar Atmos. Symp. Pt. II* (London: Pergamon Press), p. 165.
 GREENHOW, J. S., and NEUFELD, E. L., 1959, *Proc. phys. Soc. Lond.*, **74**, 1.
 HINES, C. O., 1955, *J. atmos. terr. Phys.*, **7**, 14; 1956, *Ibid.*, **9**, 56; 1959, *J. geophys. Res.*, **64**, 939.
 HOWELLS, I. D., 1960, *J. Fluid Mech.* (to be published).
 JONES, I. L., 1958, *J. atmos. terr. Phys.*, **12**, 68.
Journal of Geophysical Research, 1959, International Symposium on Fluid Mechanics of Ionosphere.
 MAEDA, K., KATO, S., and TSUDA, T., 1959, *J. Geomagn. Geoelect.*, **10**, 91.
 MOYAL, J. E., 1952, *Proc. Camb. phil. Soc.*, **48**, 329.
 MÜNCH, G., and WHEELON, A. D., 1958, *Physics of Fluids*, **1**, 462.
 RATCLIFFE, J. A., 1956, *Rep. Progr. Phys.*, **19**, 188.
 SPITZER, L., 1956, *Physics of Fully Ionized Gases* (New York: Interscience).
 VILLARS, F., and WEISSKOPF, V. F., 1955, *Proc. Inst. Radio Engrs, N.Y.*, **43**, 1232.
 WHEELON, A. D., 1957, *Phys. Rev.*, **105**, 1706.

Direct Observation of Defects in Evaporated Silver†

By V. A. PHILLIPS

General Electric Research Laboratory, Schenectady, New York

[Received March 10, 1960]

ABSTRACT

Single-crystal films of silver were made by evaporation of silver onto the cleavage surface of a heated rock-salt crystal. The crystalline defects present in the detached films were studied by transmission electron microscopy. In fresh areas of the foil, numerous simple stacking faults confined to a single plane were observed. The number varied between 5×10^9 and $1 \times 10^{10} \text{ cm}^{-2}$. There was a strong tendency for the partial dislocations to lie in $\langle 110 \rangle$ directions. The morphology of the simple fault shapes resulting is considered in detail. Electron beam stressing resulted in the dissociation of whole dislocations and in the widening of faults already present. The interaction of the faults to form complex faults on several planes, and the effects of holes and dislocations were studied. Some areas showed minor defects consisting of dislocation loops and tetrahedral stacking faults. Preliminary annealing experiments indicate that the simple faults are remarkably stable at 300°C . The origin of the various defects is discussed. It is concluded that most simple faults probably form by dissociation of dislocations which may (1) grow in from the substrate, (2) nucleate at silver patch edges during growth, (3) form by impingement of patches on a $\{110\}$ plane during growth.

§ 1. INTRODUCTION

UNTIL recently the electron microscope had not been used appreciably to study the structure of metal films condensed onto single-crystal substrates. It is, of course, well known that single-crystal films of metals may be made in this way as a result of 'epitaxy'. Pashley (1956) has reviewed the methods of achieving this. As a result of the work of Bollmann (1956 a, b), Hirsch *et al.* (1956), Whelan and Hirsch (1957 a, b) and Whelan *et al.* (1957), techniques are available for directly observing crystalline defects such as whole and partial dislocations and stacking faults.

The present work is an application of the direct observation techniques to the study of defects present in evaporated films of silver, both as-grown and after stressing in the beam of the electron microscope. This work was carried out independently of the related study of Matthews (1959) on evaporated films of gold, silver and copper, and does not overlap greatly. Pashley (1959) and Bassett and Pashley (1958-59) have investigated the structure of evaporated films, particularly of gold, but their results are complementary to the present work. A preliminary account of the present work has been given (Phillips 1959).

† Communicated by the Author.

§ 2. EXPERIMENTAL DETAILS

Films were prepared by evaporation of silver of 99.999% purity from a conical tungsten filament onto a freshly cleaved rock-salt crystal maintained at 300°C. The evaporation was carried out in a vacuum of about 10^{-4} mm of mercury. The films were cooled on the substrate *in vacuo* to room temperature which took about 20 min. The silver was floated off on a water surface, washed and pieces transferred to specimen grids.

Single crystal foils with an $\{001\}$ plane parallel to the surface were obtained. The thickness of most foils ranged from 400–700 Å, although in the immediate vicinity of holes they were, of course, thinner.

Specimens were examined in a Siemens Elmiskop I electron microscope operated at 100 kv. Double condenser illumination was used. The samples were tilted by means of a stereo-holder to bring any defects present into contrast. A ciné film was made of the effects observable at about $\times 40\,000$ on the fluorescent screen. The improved resolution obtained by recording the image directly on plates was necessary for detailed study.

§ 3. IMAGING METHODS

Stacking faults in $\{001\}$ films are readily seen by diffraction contrast and typically show extinction fringes parallel to the intersection of the $\{111\}$ fault plane with the foil surfaces, that is, parallel to $\langle 110 \rangle$ directions in the plane of the film. The detailed theory of these fringe effects has been worked out by Whelan and Hirsch (1957 a, b). Figure 1 (schematic) depicts the type of electron diffraction pattern given by an $\{001\}$ silver film. The 'crosses' such as that enclosed by circle B are a well-known feature of such patterns. (See for example Watanabe 1956, Ogawa *et al.* 1955.)

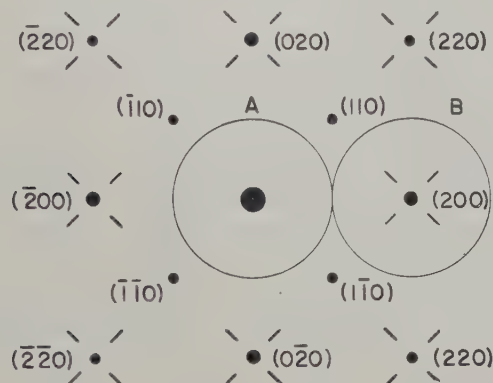
In the present work stacking faults were examined both in bright field and in dark field. The respective objective aperture positions for these two methods are shown by the circles A and B in fig. 1. The objective aperture size was $50\,\mu$. Dark field images were obtained using the portion of the diffraction pattern encircled at B comprising a $\{200\}$ reflection and satellite spots. The gun was tilted to bring an $\{002\}$ onto the axis of the microscope. Under these conditions, individual faults contributing to this portion of the diffraction pattern are brightly imaged and individual fringes are not usually discernible. Individual dislocations may be seen in dark contrast where there is an overlapping stacking fault. Improved resolution of individual faults at moderate magnifications appears to be obtained in dark field. In bright field individual faults may be imaged in dark contrast without resolving individual fringes. This usually occurs in the vicinity of a diffuse tilt fringe as in fig. 2† and some portions of fig. 4(b). The stacking fault contrast here is probably dynamical in origin. More commonly, in bright field, individual fringes are resolved at stacking faults in the way predicted by theory (Whelan 1958–59).

† Figures 2 to 9 are shown as plates.

§ 4. STRUCTURE OF THE FILMS

Figures 2-9 are electron micrographs of the films. Stacking faults were observed in all areas. The number of stacking faults varied from one film to another but was usually between 5×10^9 and 1×10^{10} per cm^2 . This number is relatively independent of the thickness since most faults ran from surface to surface. There was a big difference in the length and complexity of the faults depending on whether or not a particular area of foil had been previously exposed to the electron beam. Exposure to the beam caused buckling, as evidenced by movement of tilt fringes across the field, and the stresses produced caused wide dissociation and frequent combination, and interaction of stacking faults both with each other and with dislocations and holes through the foil. These effects will be discussed later in more detail. Figures 2 and 4 are typical of the structure observed in areas of foil not previously irradiated. Figure 5 is typical of the appearance after exposing to the beam for some time.

Fig. 1



Typical electron diffraction pattern from a silver film (schematic). Circles A and B show the portions of the diffraction pattern used in the light and dark field imaging methods.

Defects which appear to be typical dislocation loops were observed in fresh areas of the films as at A in fig. 2. These loops ranged in diameter from about 100-300 Å. Smaller loops would not be recognizable but may well be present. The average number of loops in the region shown, which is about 650 Å thick, is approximately $2 \times 10^{14} \text{ cm}^{-3}$. Assuming an average loop diameter of 200 Å, in order to calculate the total line length per unit volume, ρ , this loop number is equivalent to a dislocation density of $\rho/3$, that is, $4 \times 10^8 \text{ cm}^{-2}$. Loops of similar sizes have been observed in quenched aluminium by Hirsch *et al.* (1958). If the loop is assumed to be a collapsed disk of vacancies, then for an average diameter of 200 Å, the corresponding initial concentration of vacancies is 1.5×10^{-5} . Dislocation loops have been observed in quenched bulk silver by Smallman and Westmacott (1959).

A number of parallel line features will be noted in fig. 2, for example B_{1-3} . These are not due to slip contrast since this appeared to fade in less than 1/10 sec. They are instead attributed to movements of loops during the 3 sec period of the exposure. The projected direction of movement is parallel to a $\langle 110 \rangle$ direction in the plane of the film. Loops B_1 and B_2 (fig. 2) appear to have slipped out to holes through the foil. B_3 appears to have moved in a jerky fashion and is still visible. If this interpretation is correct, it implies that the loops are prismatic loops, formed by nucleating a Shockley partial in the disk of stacking fault associated with the Frank sessile produced by collapse of a disk of vacancies on a $\{111\}$ plane (Kuhlmann-Wilsdorf 1958). These would glide on a cylinder whose axis is parallel to a $\langle 110 \rangle$ direction not lying in the plane of the fault. Opposite halves of the loop would presumably glide in opposite directions as observed by Hirsch *et al.* (1958) in aluminium. Some doubt is thrown on this interpretation by the fact that the stacking-fault energy in silver is estimated (Thornton and Hirsch 1958) to be only about 35 erg cm^{-2} which is hardly sufficient to make the nucleation of a Shockley partial energetically favourable unless stress-assisted. The glissile nature of the loops would explain why they are only observed on first examination of an area, and why in similar observations on evaporated gold (Bassett and Pashley 1958-59), they apparently vanished between exposures.

Perfect dislocations were commonly seen running obliquely through the film as at D in fig. 3. Dislocations running perpendicularly to the film from surface to surface would appear as black dots in bright field as for example at C in fig. 3. Many dot-like features are also visible in fig. 6, some of which are superimposed on stacking faults. Since there is no interaction with the faults they cannot be line dislocations running right through the film but must be small defects away from the plane of the fault, possibly small loops of less than 50 \AA diameter formed by vacancy condensation.

Tetrahedral stacking faults were occasionally observed confirming the finding of Matthews (1959). Smallman and Westmacott (1959) claim to have observed tetrahedra in quenched and aged bulk silver. Examples of tetrahedra are visible at E in fig. 3: the largest has a side length of about 500 \AA , and the smallest about 150 \AA . This agrees well with the range of sizes of tetrahedra observed by Silcox and Hirsch (1959) in quenched gold, which is to be expected since the stacking fault energy of gold and silver are about the same.

Small triangular defects such as F in fig. 3 and G in fig. 4(a) were quite commonly observed. These correspond to equilateral triangles in a $\{111\}$ plane, the edges of which are parallel to $\langle 110 \rangle$ directions lying in that plane. The smallest defects recognizable as of this type have side lengths of about 100 \AA . The largest would have sizes limited by the thickness of the film; for example, that at A' in fig. 4(a) has a side length of about 650 \AA .

The evidence is compatible with the view that the small triangular defects which do not in general intersect the film surface are Frank sessile

dislocations (Frank 1949) formed by a vacancy condensation mechanism, whereas the large defects which generally intersect the film surfaces and of which A' (fig. 4(a)) is one sort are of a different type. These large defects will be discussed in §§ 5 and 6. The triangular shape of the small defects is expected, since Kuhlman-Wilsdorf (1958) has pointed out that a Frank sessile dislocation will prefer to split up into segments parallel to $\langle 110 \rangle$ directions.

§ 5. DESCRIPTION OF SIMPLE STACKING FAULTS

In areas of the films not previously exposed to the beam most of the faults were relatively short and confined to a single plane as in figs. 2 and 4. Faults confined to a single plane will be called 'simple faults'. Complex faults forming on two or more of the four possible $\{111\}$ planes were usually only observed after irradiation in the beam. Matthews (1959) mainly observed complex faults and has described these rather fully. The nature of simple faults will now be described.

Most of the simple faults observed, with the exception of those ending at a hole, were straight-sided with 3–6 sides and extended from surface to surface of the film. About 5% of the partials assumed the shortest possible length parallel to a $\langle 112 \rangle$ direction so that the fault viewed in a direction perpendicular to the foil appeared 'square-ended'. A strong tendency was observed for the partial dislocations to align themselves in $\langle 110 \rangle$ directions as predicted by Kuhlmann-Wilsdorf (1958). Analysis of the directions of about 50 partials in figs. 2 and 3 showed that about 60% of them coincided with a possible $[110]$ direction within the accuracy of measurement, and about 90% were within 15° . The direction of significant deviations was always such as to shorten the length of the partial.

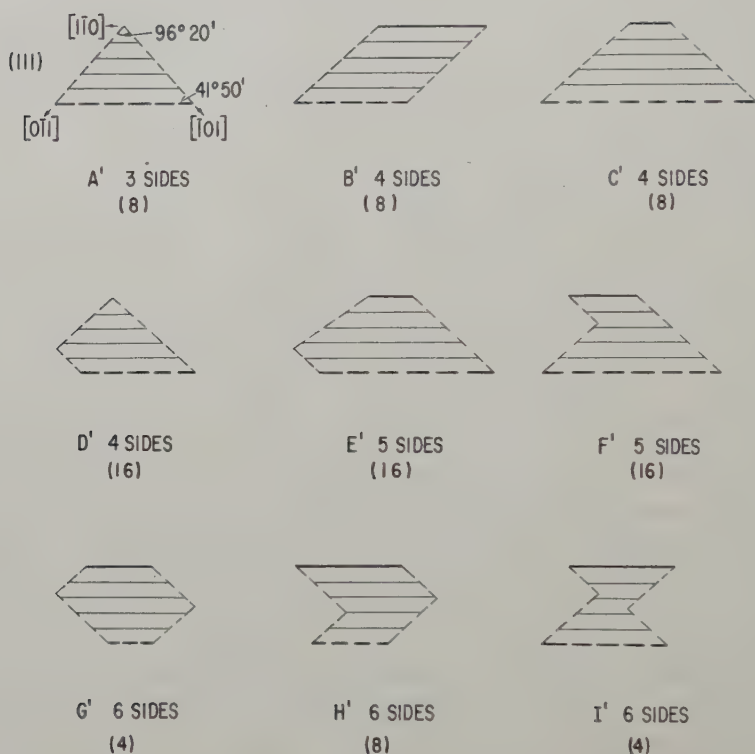
The stacking faults lie on any of the four $\{111\}$ planes making a $54^\circ 44'$ angle with the plane of the foil. The latter will be taken as (001), so that the (111) and $(11\bar{1})$ planes intersect the surface along $[1\bar{1}0]$ while the $(1\bar{1}1)$ and $(\bar{1}11)$ planes intersect it along $[110]$. It is not possible from the present micrographs to determine which of the fault boundaries is in the top surface of the foil, so that the stacking fault at T' in fig. 2, for example, is either on the (111) or on the $(11\bar{1})$ plane if the long edge of the figure is parallel to $[1\bar{1}0]$.

Line drawings of the main possible shapes of simple fault projected onto the plane of the film are shown in figs. 10–13. The shapes have been classified according to the number of sides and the side direction. The convention adopted by Matthews (1959) is followed, viz. the intersections of the fault with the top and bottom surfaces of the film are shown by thick full and dashed lines, respectively. Fine full lines represent the extinction fringes caused by the fault. Dashed fine lines represent partial dislocations. In each of the four $\{111\}$ fault planes there are three possible $\langle 110 \rangle$ directions, one of which lies in the plane of the foil. Thus in the (111) plane, the partial may be in the $[0\bar{1}1]$ or $[\bar{1}01]$ directions or may take the shortest distance through the film in the $[\bar{1}\bar{1}2]$ direction. The intercept of the (111) plane with the surfaces and also the extinction fringe direction are

parallel to $[1\bar{1}0]$. The projection of either $[0\bar{1}1]$ or $[101]$ perpendicular to the foil makes an angle of 45° with the $[110]$ direction, while the projection of $[\bar{1}\bar{1}2]$ is at 90° to $[1\bar{1}0]$.

Faults of type A' may occur in two ways in each of the four planes, that is in eight ways. Similar information for all the fault types is given in brackets in figs. 10–13. Since the top and bottom fault boundaries are indistinguishable in the micrographs only a half of these possibilities would

Fig. 10



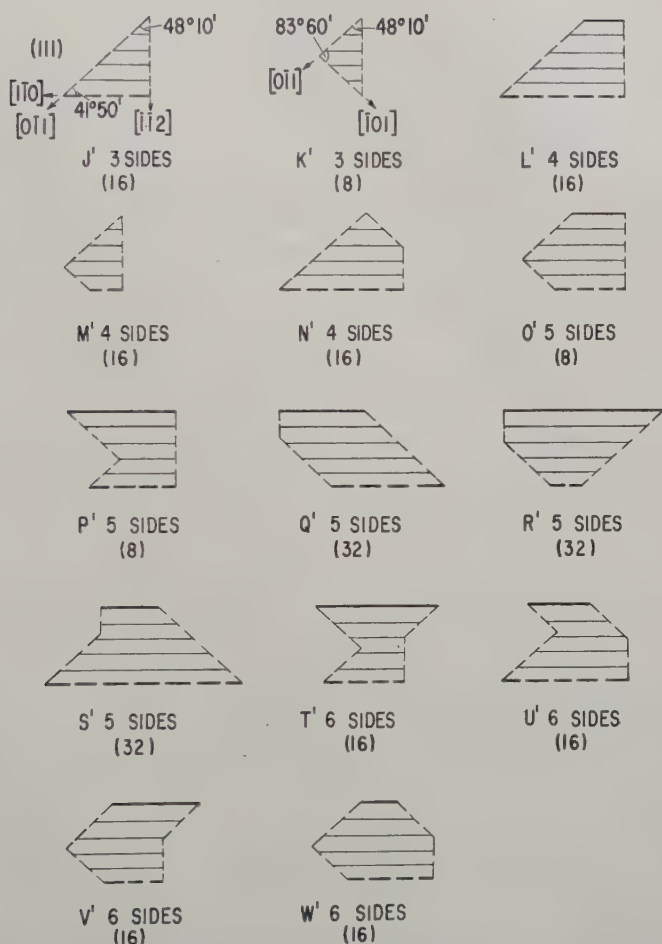
Possible shapes of simple fault with partials in $\langle 110 \rangle$ directions.

be distinguishable. Examples of 19 types of fault are identified in fig. 4 (a) and may be seen in fig. 4 (b) in light field. Examples of three more types are shown in fig. 2. Examples of G', M', N', O', P', V', W', AB', AC' and AF' are not shown. Of these, G' and AB' have been seen in other areas. The remaining types would not be expected to be common because partials parallel to $[\bar{1}\bar{1}2]$ or its equivalent are relatively infrequent. Since the partials tend to deviate somewhat from $\langle 110 \rangle$ directions, many of the shapes deviate somewhat from those described.

Although there was a strong tendency for the partials to be straight, they were occasionally curved in a manner more typical of stainless steel (Whelan *et al.* 1957) as at fault T in fig. 6(c).

Many of these simple fault types are also to be seen in a micrograph of gold grown directly on rock-salt published by Bassett and Pashley (1958-59). It is probable that they are typical of face centred cubic metals of low stacking fault energy grown epitaxially on rock-salt.

Fig. 11

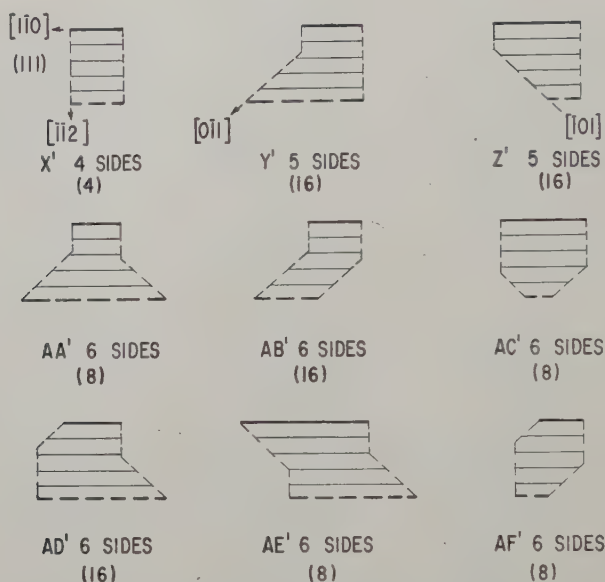


Possible shapes of simple fault with one partial in $\langle 110 \rangle$ directions, other in $\langle 112 \rangle$ or $\langle 110 \rangle$ directions.

Comparison of fig. 4(a) with 4(b) shows that most, but not all faults are in contrast in fig. 4(a). Some changes have occurred between the two exposures, for example at L and M. These will be discussed in § 6. The object I in fig. 4 is probably a twin. This is one of the very few examples observed.

H_1 in fig. 4(a) is an example of a group of faults on parallel planes spaced 200–500 Å apart. H_2 appears to be a similar group of three faults more closely spaced. J_1 in fig. 4(a) appears at first sight to be a similar group containing three faults, one of which is annular. However, in bright field (fig. 4(b)) the contrast is such that this interpretation appears unlikely. J_2 in fig. 4(a) and (b) is a similar example. The nature of these defects is uncertain.

Fig. 12



Possible shapes of simple faults with both partials in $\langle 112 \rangle$ or partly in $\langle 112 \rangle$ and partly in $\langle 110 \rangle$ directions.

The smallest defects such as K imaged in fig. 4(a) are about 50 Å in diameter. Comparison with fig. 4(b) shows that most of them are located at the edges of holes. These are thought to be tiny faults as discussed in § 6.

It is possible that simple planar defects are common in bulk material but have not yet been recognized. Thus the triangular defects observed in quenched and aged bulk silver by Smallman and Westmacott (1959) and Smallman *et al.* (1959–60) appear to be of type A' although interpreted by the authors as defect tetrahedra. Similarly, some of the defects observed by Silcox and Hirsch (1959) in quenched and aged gold could be simple planar faults of types A' and C' rather than tetrahedra.

§ 6. DISSOCIATION AND FORMATION OF COMPLEX FAULTS

Whole dislocations lying in $\{111\}$ planes tended to dissociate into two (Shockley) partials parallel to $\langle 110 \rangle$ directions and connected by stacking fault. Dislocations Q_1 and Q_2 in fig. 6 illustrate this and also show how

this dissociation tends to be inhibited by the stress field associated with the dislocations R_{1-4} piled-up on a neighbouring parallel plane against a group of stacking faults. The maximum dissociation of Q_2 is about 200 \AA in (a), 100 \AA in (b) and 450 \AA in (c). The dissociation of Q_1 in fig. 6(b) has produced a defect of type C' (fig. 10). Other examples of the blocking of the movement of whole dislocations by stacking faults are seen at Z in fig. 6 and at Y where the dislocation has moved out of the field between the exposures in fig. 6(b) and (c).

Some individual dislocations such as D in fig. 3 showed no resolvable dissociation either because they did not lie in a $\{111\}$ plane or because dissociation was inhibited by the local stress field as at R_3 in fig. 6(b). Dissociation to a width of less than about 100 \AA would not be detectable.

Stressing in the beam caused the short simple faults typically seen in figs. 2 and 4 to widen and interact. Holes, dislocations and other faults acted as barriers to dissociation. Eventually most faults have extended as far as possible as in fig. 5 and dislocation movements become less frequent. For an average (extended) fault width of say 0.3μ , and assuming 10^{10} faults/cm², an order of magnitude calculation shows that this corresponds to a plastic strain of about 5×10^{-8} . Figure 6 illustrates the structure at an intermediate stage.

Dislocation movements were very readily produced in the foil. This is attributed partly to the cleanliness of the surface, for it is probable that oxide or other contaminant layers tend to impede dislocation movement, and partly to the presence of holes which acted as stress concentrators. Holes also acted as pinning points for extended faults as at P in fig. 6(a) where one partial has run out to the hole surface giving a fault of type AG' in fig. 13. Other fault shapes such as AI' in fig. 13 may be observed. Exceptionally a fault may run into two holes eliminating both partials giving a fault of type AH' in fig. 13. Thin places in the vicinity of holes or elsewhere can similarly reduce the length of both partials and stair-rod dislocations. Movement of such partial dislocations would cause lengthening and is thus made difficult. Effects of this kind may stabilize normally unstable configurations of faults.

The widening of a fault under stress occurs when one partial is pinned and the other moved under the action of the stress. Since movement of a partial produces a shear, the stress may be locally relieved and movement stops, however, the shear produces a redistribution of the complex local stresses so that another fault is affected and expansion or contraction of this fault occurs. Thus movements tended to be rapid but to occur in stages, for example, fault S in fig. 6(a) has extended a short distance in fig. 6(b) and in fig. 6(c) has extended to a hole where the partial will tend to be anchored due to the reduction in its length.

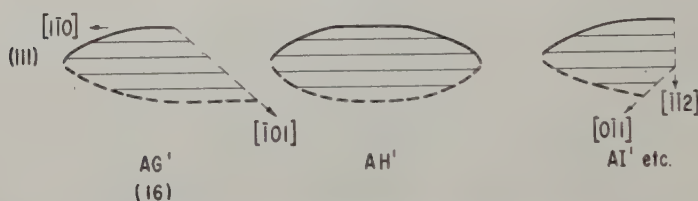
Combination reactions leading to complex stacking faults have been discussed by a number of authors (Cottrell 1952, Thompson 1953, Frank and Nicholas 1953, Friedel 1955, Whelan 1959). The Burgers vectors of the stable stair-rod dislocations have been listed by Frank and Nicholas

(1953). While such considerations are probably applicable to fully annealed clean films of uniform thickness, they are of doubtful value here because it is clear that the reduction in line energy of both partial and stair-rod dislocations at holes and thin places plays an important role, as also does the blocking of the movement of partials by other faults and dislocations. The formation of complex faults as a result of local stressing in the electron beam will now be considered.

A common reaction observed was the combination of two simple faults to form a one-bend fault. This has occurred at L and M between the consecutive exposures shown in fig. 4(a) and (b). The angle between the faulted planes at L is acute ($70^{\circ} 32'$) and at M is obtuse ($109^{\circ} 28'$). The stair-rod dislocations (Thompson 1953) at the bends cannot here be resolved. Examples of stair-rods may be seen in fig. 6(a) at the obtuse angled fault S and at the acute angled fault γ . Partial dislocations are sometimes visible, thus both partials are visible in fig. 6(b) at the single bend fault W.

Another example of the formation of a single-bend acute-angled fault from two simple faults is shown in a ciné sequence (fig. 7(a) to (c)).

Fig. 13



Possible shapes of simple faults running into holes—one or no partials.

The fact that the triangular fault at L in fig. 4(a) combined with the other fault would suggest that it is bounded by Shockley rather than Frank partials lending some support to the distinction of two types of triangular fault in § 5.

The nucleation of dislocations at thin foil edges in stainless steel under the action of thermal stresses has been observed by Whelan *et al.* (1957). They calculate, following a private communication from Cottrell, that it is relatively easy for a dislocation to nucleate at a thin edge if it is of edge type. It was commonly observed in silver that faults tended to nucleate from the edges of holes through the films. The tiny faults K (fig. 4(a)) which are located at hole edges (fig. 4(b)) are thought to represent a very early stage in this process. These faults are only 50–100 Å in diameter. Figure 6(a) to (c) shows the apparent nucleation of three successive faults from the hole edge at point W. The first and last faults to form are on parallel planes as indicated by the change in fringe contrast. A further example of simple overlapping faults is seen, at X in fig. 6(b); these are seen to have

lengthened in fig. 6(c). Another instance of fault nucleation at a hole is shown at N in fig. 9. The source of the dislocations Q_1 and Q_2 in fig. 6 is probably the hole edge to the left of Q_1 .

The behaviour is frequently more complex. Comparison of fig. 8(a) and (b) shows that the single-bend fault DEF has become a complex double-bend fault DEHGFI.

Simple faults were frequently observed to become shorter as at PQ in fig. 9(b) to (c). Complex faults were frequently observed to contract forming simpler faults. Examples of this behaviour are seen at UT in fig. 6(a) to (c) and at JKLM in fig. 9(a) to (c).

Preliminary experiments have been made in which films were annealed *in vacuo* on grids for half an hour at 300°C followed by slow cooling. It appears that this treatment reduced the number of faults by a factor of 2-3. After annealing most of the faults observed are short simple faults. Whole dislocations are also present. Due to a loss in resolution occasioned by surface deterioration during annealing it was not possible to study the minor defects. Electron beam stressing caused the simple faults to widen and interact to form complex faults as before. This annealing experiment emphasizes the remarkable stability of the simple faults.

Widening and interaction processes of the kind described create complex fault patterns which impede free dislocation movement. These processes may play an important role in work hardening in bulk materials such as silver and gold if simple faults are present which may be the case as suggested in § 6. If such defects are peculiar to evaporated films then they may account for the claims that the strengths of such films may considerably exceed that of bulk material.

§ 7. ORIGIN OF INITIAL DEFECTS

The formation of the minor defects present namely dislocation loops, stacking-fault tetrahedra and small triangular defects is attributed to a vacancy collapse mechanism as discussed in § 4. The dot-like features are believed to be some kind of vacancy aggregate, perhaps unresolved loops. A vacancy concentration of about 5×10^{-5} is sufficient to account for the formation of all these defects. This seems a reasonable value for an evaporated foil.

The major defects present, namely stacking faults with their associated dislocations, could have formed by a number of alternative processes:

- I. Matthews (1959) growth impingement mechanism.
- II. Vacancy collapse mechanism.
- III. Growth mistakes.
- IV. Dissociation of whole dislocations.

These will now be considered in turn.

Matthews supposes that adjoining nuclei of silver are of different heights, grow with $\{111\}$ facets, and impinge on a $\{111\}$ plane with a mis-match in the stacking sequence. By suitable elaboration, the formation of both

simple and complex faults can be explained. If this mechanism were important, faceting would be evident at the edges of the holes through the present films. This is not observed. A further objection to Matthews mechanism is that growth of nuclei in the case of evaporated gold (Bassett and Pashley 1958-59) and presumably in silver, seems in the early stages to occur by the growth of some nuclei at the expense of others rather than by impingement and coalescence, yet stacking faults appear to form at a very early stage in the growth before impingement becomes common.

A vacancy collapse mechanism will now be considered. In a fresh area of film, such as that shown in fig. 4, the density of major faults is about $1 \times 10^{10} \text{ cm}^{-2}$. Assuming an average width of 0.1μ which is fairly typical of a fresh area and assuming that the faults formed by vacancy aggregation and collapse, this density is equivalent to an initial vacancy concentration of about 3×10^{-3} . This seems unreasonably high. Furthermore, most of the partial dislocations appear to be glissile in the plane of the associated fault. These considerations appear to rule out a vacancy collapse mechanism.

Growth mistakes could occur at an early stage during growth, possibly in laying down the first few $\{111\}$ layers, which could account for the presence of faults running from surface to surface of the films. The low stacking fault energy in silver would increase the probability of such mistakes occurring. More evidence is needed in order to properly assess the importance of this mechanism.

The most important mechanism of formation of stacking faults is thought to be the dislocation dissociation one. Dislocations might be nucleated at silver patch edges during growth since it is probable that the silver is in a state of stress. Secondly, they could form by the impingement of adjacent patches on a $\{110\}$ plane during the later stages of growth. Thirdly, it is also likely that dislocations grow from the substrate. Thus edge dislocations with a $\langle 110 \rangle$ Burgers vector emerging on a $\{100\}$ cleavage surface of the rocksalt could grow into the silver during the first stage of deposition, and as growth continued would tend to dissociate into the appropriate Shockley partials. Since there is a marked tendency for the partials to lie in $\langle 110 \rangle$ directions, it would be expected that faults formed in these ways would commonly tend to be triangular in shape (type A'). The driving force for movement of the partials into $\langle 110 \rangle$ directions presumably comes from a reduction of core energy in this orientation. Since the stacking fault energy is low, only a small stress is then required to cause further broadening of type A' faults to form type C'. Considerable stresses arise during cooling of the film on the substrate due to differential contraction and also during stripping and drying due to surface tension effects. If both partials lay in the same $\langle 110 \rangle$ direction a fault of type B' would be expected and could readily widen under stress. Faults such as type AG' could result from the nucleation of a single Shockley partial at a patch edge. As demonstrated, further dissociation and interaction to form complex faults can readily occur. New faults can also be nucleated.

ACKNOWLEDGMENTS

The author is indebted to Miss E. Alessandrini who prepared the silver films and to Mr. J. A. Hugo for his assistance in the experimental work. He would like to thank his colleagues, Dr. W. G. Johnston and Dr. J. D. Livingston for criticizing the manuscript.

REFERENCES

- BASSETT, G. A., and PASHLEY, D. W., 1958-59, *J. Inst. Met.*, **87**, 449.
 BOLLMANN, W., 1956 a, *Phys. Rev.*, **103**, 1588; 1956 b, Report of First European Regional Conference on Electron Microscopy, Stockholm.
 COTTRELL, A. H., 1952, *Phil. Mag.*, **43**, 645.
 FRANK, F. C., 1949, *Proc. phys. Soc. Lond. A*, **62**, 202; 1950, *Symposium on the Plastic Deformation of Crystalline Solids*, O.N.R. (Pittsburgh: Carnegie Institute of Technology), p. 150.
 FRANK, F. C., and NICHOLAS, J. F., 1953, *Phil. Mag.*, **44**, 1213.
 FRIEDEL, J., 1955, *Phil. Mag.*, **46**, 1169.
 HIRSCH, P. B., HORNE, R. W., and WHELAN, M. J., 1956, *Phil. Mag.*, **1**, 677.
 HIRSCH, P. B., SILCOX, J., SMALLMAN, R. E., and WESTMACOTT, K. H., 1958, *Phil. Mag.*, **3**, 897.
 KUHLMANN-WILSDORF, D., 1958, *Phil. Mag.*, **3**, 125.
 MATTHEWS, J. W., 1959, *Phil. Mag.*, **4**, 1017.
 OGAWA, S., WATANABE, D., and FUJITA, F. E., 1955, *J. Phys. Soc. Japan*, **10**, 429.
 PASHLEY, D. W., 1956, *Advanc. Phys.*, **5**, 173; 1957, Proceedings of the Conference on Electron Microscopy held at Bangor, Wales (summarized by EMERSON, H. W., 1958, *Brit. J. appl. Phys.*, **9**, 306); 1959, *Phil. Mag.*, **4**, 324.
 PHILLIPS, V. A., 1959, Annual Meeting of the Electron Microscope Society of America held at Ohio State University, Columbus, Ohio (abstracted in *J. appl. Phys.*, 1959, **30**, 2032).
 SILCOX, J., and HIRSCH, P. B., 1959, *Phil. Mag.*, **4**, 72.
 SMALLMAN, R. E., and WESTMACOTT, K. H., 1959, *J. appl. Phys.*, **30**, 603.
 SMALLMAN, R. E., WESTMACOTT, K. H., and COILEY, J. H., 1959-60, *J. Inst. Met.*, **88**, 127.
 THOMPSON, N., 1953, *Proc. phys. Soc. Lond.*, B, **66**, 481.
 THORNTON, P. R., and HIRSCH, P. B., 1958, *Phil. Mag.*, **3**, 738.
 WATANABE, Y., 1956, *J. Phys. Soc., Japan*, **11**, 740.
 WHELAN, M. J., 1959, *Proc. roy. Soc. A*, **249**, 114.
 WHELAN, M. J., and HIRSCH, P. B., 1957 a, *Phil. Mag.*, **2**, 1121; 1957 b, *Ibid.*, **2**, 1303.
 WHELAN, M. J., HIRSCH, P. B., HORNE, R. W., and BOLLMANN, W., 1957, *Proc. roy. Soc. A*, **240**, 524.

Atmospheric Temperature Effects on the Solar Daily Variation of Cosmic Ray Intensity†

By J. J. QUENBY and T. THAMBYAHPIILLAI
Imperial College of Science and Technology, London

[Received January 20, 1960]

ABSTRACT

The solar diurnal variation of the cosmic ray meson intensity caused by the periodic solar heating of the atmosphere is derived from a comparison of the ionization chamber and neutron monitor data from Huancayo. The amplitude and the time of maximum are found to be 0.11% and 05–30 hr respectively. Recent meteorological data, free from appreciable radiation errors, are found to be in rough agreement with this result. Particularly good agreement is obtained in the phases which are both later than the value given by Dorman by about 4 hr. The inclusion of this temperature correction enables the directional measurements at Mawson to be reconciled with the idea of a primary anisotropy.

§ 1. INTRODUCTION

THE solar daily variation of cosmic ray intensity has been the subject of extensive investigation, chiefly because of the important information it can yield about the anisotropic distributions of the primary radiation and the modulating mechanisms responsible for their production. Continuous monitoring of the intensity, however, has been feasible only with the secondary radiation at sea-level or mountain altitudes where measurements have been made with ionization chambers, counter telescopes and neutron monitors. The intensity variations of the secondary radiation, however, are governed not only by the primary intensity variations but also by the variable conditions of the overlying atmosphere, and adequate corrections have to be applied for such atmospheric influences in order to relate the observed variations to those in the primary intensity.

The neutron-producing component of the secondary radiation is sensitive to changes of barometric pressure but is not affected to any appreciable extent by atmospheric temperature. Since effects of barometric changes can be taken adequately into account, observations of the neutron intensity provide reliable measures of the solar diurnal variations of the primary radiation entering the earth's atmosphere. In contrast, the total ionizing and μ -meson components measured by ionization chambers and counter telescopes are influenced both by pressure variations and variations of temperature throughout the atmosphere. As will be seen later, difficulties are encountered in attempting to correct for the atmospheric temperature effects.

† Communicated by the Authors.

In spite of these difficulties over meteorological effects, counter telescopes possess an advantage over neutron monitors in their ability to select the radiation incident from particular directions and because of this, are of particular importance in a determination of the true direction and magnitude of the anisotropy existing outside the earth's magnetic field. For instance, Brunberg and Dattner (1954) have shown how, by making some simple assumptions, the mean energy of the primary radiation giving rise to the solar daily variations and the direction of the anisotropy relative to the sun-earth line can be deduced from the directional measurements made with telescopes inclined at 30° to the vertical and pointing in the north and south directions.

Measurements of this kind made by Elliot and Dolbear (1951) and Malmfors (1949) have shown that the observed solar daily variations were of about the same magnitudes as those originating in atmospheric temperature variations. Under these circumstances, it is necessary to know the meteorological corrections with considerable accuracy in order to obtain any reliable information concerning the anisotropy.

Now, although many elegant theoretical investigations have already contributed to an understanding of atmospheric effects on the μ -meson intensity near sea-level (Olbert 1953, Maeda and Wada 1954, Trefall 1955, Dorman 1957), corrections for meteorological influences on the solar daily variations of meson intensity have not hitherto been evaluated to a high degree of accuracy. The chief difficulties encountered have been a dearth of aerological information and where radio-sonde measurements of atmospheric temperature were available, the uncertainty as to the extent to which the measurements were contaminated by insolation and lag errors.

There is general agreement that the radio-sonde measurements of the diurnal variation at high altitudes do not reflect the actual diurnal course of the air temperature (Kay 1951, Dorman 1957). However, Dorman has estimated the atmospheric contribution to the diurnal variation of meson intensity on the assumption that the radio-sonde measurements were reliable up to the 300 mb pressure level. The diurnal amplitude in the region above this isobar was estimated from cosmic ray measurements made underground at London to be $< 0.5^\circ\text{C}$ while the upper air soundings gave amplitudes of over 2°C . When errors of over 300% occur just above 300 mb, the limit of integration, it is conceivable that substantial radiation errors are also present in the measurements made below 300 mb and to this extent, the estimate of Dorman appears to be unreliable. An alternative estimate of the atmospheric contribution is therefore desirable.

In the following, we have attempted to deduce the 24 hr wave in meson intensity due to atmospheric temperature using only cosmic ray measurements and compare this result with the most recent meteorological data. Finally, it is shown that when this temperature correction is included, the directional measurements made at Mawson can be explained in terms of a primary anisotropy without any inconsistencies.

§ 2. COMPARISON OF NEUTRON MONITOR AND IONIZATION CHAMBER DATA FROM HUANCAYO

In this section the diurnal variation of meson intensity arising from changes of atmospheric temperature and hereafter referred to as the temperature wave, will be derived from a comparison of the ionization chamber and neutron monitor data from the equatorial station Huancayo. In this particular case where the response of the two detectors to primary intensity variations is likely to be similar, the basic cause of the difference between the two sets of results, after correction for barometric changes, is the residual atmospheric contribution left in the ion-chamber data.

We shall first examine the basis for the assumption of similarity in detector response. Rough similarities are evident in (i) Directional Sensitivity and (ii) Energy Response as defined by the differential response curves giving the number of detected secondary particles produced, as a function of the primary magnetic rigidity. These will now be described briefly under (i) and (ii)

- (i) Both instruments are omni-directional in the detection of the secondary radiation and consequently the directional sensitivities to the primary radiation are determined solely by atmospheric absorption.
- (ii) It is evident from the differential response curves given by Webber and Quenby (1959) that at rigidities higher than the equatorial cut-off value, the curves for an ion-chamber and a neutron monitor are similar at the atmospheric depth of Huancayo.

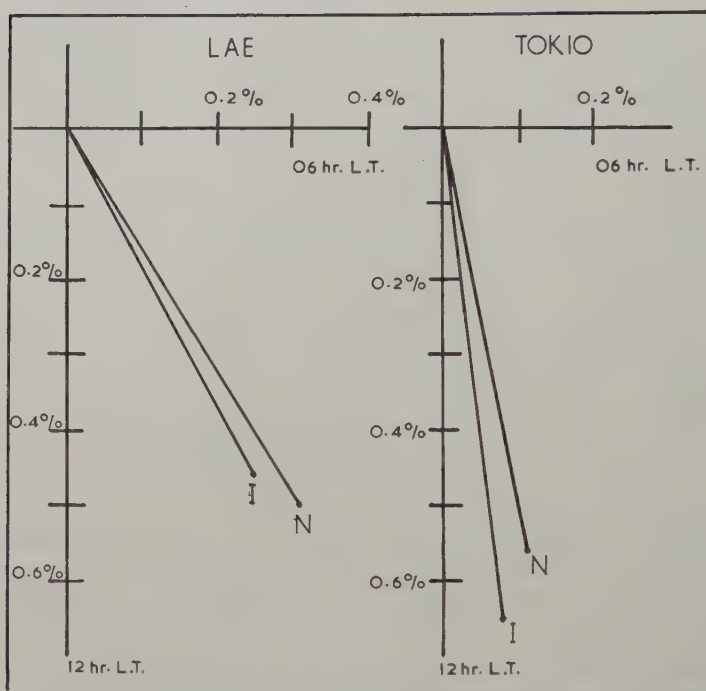
The following experimental result lends further support to the belief that the instrumental responses are similar. It is known that during certain periods, particularly after magnetic and cosmic-ray storms, the amplitude of the daily variation may be as great as 0.5%, whereas the temperature wave in the ion-chamber data is likely to be only 0.1% in amplitude. If the response of the two instruments to primary variations is similar, it follows that the difference between the daily variations measured by neutron monitors and meson recorders during these periods should be small.

Now, neutron monitors and wide-angled counter telescopes have concurrently been operated during 1957 and 1958 at the low-latitude stations Lae and Tokyo. The averaged results from several selected days of enhanced daily variation which followed Forbush decreases were subject to Fourier analysis and the 24 hr components are shown on harmonic dials in fig. 1. It is apparent from the diagram that the measured amplitudes are $\sim 0.6\%$ while the vectorial difference between the two detectors at any one station is $< \sim 0.1\%$. Although the data from a neutron monitor and a counter telescope have been compared here, this result should also apply to the data from a neutron monitor and ion chamber at a low-latitude station because the differences between the differential response curves of an ionization chamber and a counter telescope are

negligible (Webber and Quenby 1959). It is concluded therefore, that the two types of detector at Huancayo respond in similar ways to primary intensity variations. The evidence cited above appears to be sufficient as only a qualitative similarity and not an exact correspondence in instrumental response is required for our purpose.

Let the mean solar 24 hr waves during any given month originating from primary intensity changes be represented vectorially by **N** and **I** for the measurements made by the neutron monitor and the ion chamber

Fig. 1



Harmonic dials, showing the 24 hr components of the daily variation measured during days of enhanced variation. Simultaneous measurements made by a neutron monitor (**N**) and a wide-angled meson telescope (**I**) at Lae and Tokyo are shown.

respectively. It will be assumed that the similarity of instrumental response leads to the following relationships:

$$|\mathbf{N}| = r|\mathbf{I}|,$$

$$\mathbf{N} \cdot \mathbf{I} = \theta |\mathbf{N}| \cdot |\mathbf{I}|$$

where r and θ are constants even though the rigidity dependence and the direction in space of the primary anisotropy may change in time. These conditions will be satisfied best if, as implied earlier, r and θ are both close to unity.

Since the data have already been corrected for barometric changes, the diurnal waves given by the measurements should be \mathbf{N} for the neutron monitor and $\mathbf{I} + \mathbf{T}$ where \mathbf{T} is the atmospheric contribution to the meson data. Moreover, for the ion-chamber, the seasonal variations of the temperature wave are likely to be small at Huancayo which is near the geographical equator. Thus for practical purposes \mathbf{T} can be assumed constant.

If we now select a period of time during which the mean diurnal variation measured by the neutron monitor is zero, it can be shown that the average variation measured by the ionization chamber over this period is the temperature wave \mathbf{T} ; i.e. if $\sum_i \mathbf{N}_i = 0$ where i denotes the monthly period, then $\sum_i \mathbf{I}_i = 0$, provided that r and θ are constant. This is easily seen when we consider a number of vectors \mathbf{N}_i selected so that on addition they give a closed polygon. Even when the amplitudes of the vectors are changed uniformly by a factor r and each vector is rotated through the same angle $\cos^{-1} \theta$, the polygon will be closed. Therefore the mean 24 hr wave in the ion-chamber data is given by

$$\frac{1}{n} \sum_{i=1}^n (\mathbf{I}_i + \mathbf{T}) = \mathbf{T}.$$

The condition $\sum_i \mathbf{N}_i = 0$ can be realized using the data from the period 1953–1955. The neutron monitor data of Conforto and Simpson (1957) have been used in our analysis together with the results of Venkatesan and Dattner (1959) after allowing for some errors in the latter analysis. The definition adopted for the diurnal vector in the former publication, though different from the conventional one, is claimed by the authors to give equivalent results.

The selected groups of months during which the mean neutron diurnal effect was nearly zero are shown in table 1 together with the actual mean values and the corresponding temperature waves from the ion-chamber data. The coefficients a_1 and b_1 are given, where the diurnal wave is represented by

$$a_1 \cos \frac{2\pi}{24} t + b_1 \sin \frac{2\pi}{24} t.$$

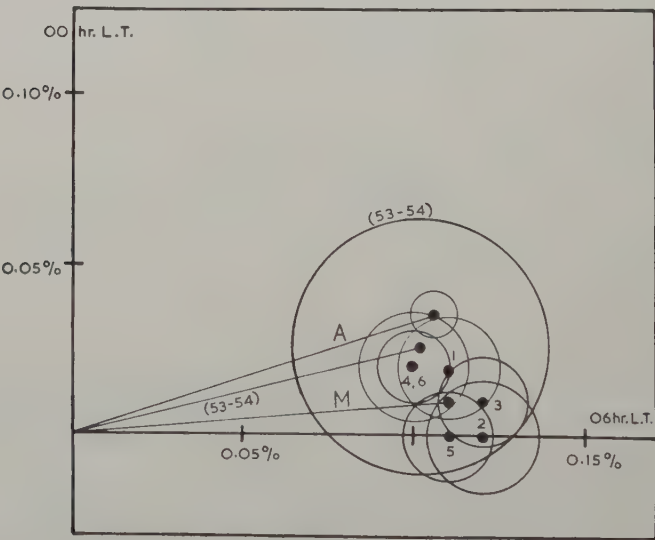
The atmospheric vectors are plotted on a harmonic dial in fig. 2 where the standard deviation circles are also drawn. The vectors for each group are numbered 1 to 6 and the overall mean is marked M . The individual vectors will be seen to agree within the limits of statistical accuracy but we note that they are not entirely independent of one another because some months are common to several groups. The mean temperature wave obtained in this way has an amplitude of 0.11% and a time of maximum around 05–40 hr local time.

Alternatively, r and θ can be determined from the data for two separate periods, again assuming r , θ and \mathbf{T} to be constant. In order to achieve accuracy in this calculation, the measured diurnal waves during the two periods should show substantial differences. The mean diurnal waves

Table 1

Group	Selected months	Mean neutron vector (%)		Atmospheric vector (%)	
		a_1	b_1	a_1	b_1
1	October, 53 Feb., March, April, May, June, July 1954	-0.002	0.01	0.02	0.11
2	March, April, September, October, November, December 54	-0.008	-0.004	0.00	0.12
3	September, October, November, December 53; January, February, March, April, August, September 54	-0.001	-0.005	0.01	0.12
4	April, May, June, July, 53, August, September 54	-0.01	+0.01	0.02	0.10
5	March, April, May, June, August September, October, November, December, 54, January, February, March, 55.	-0.004	+0.002	0.00	0.11
6	January 54 to March 55 inclusive	0.00	0.00	0.02	0.10 (+0.01, +0.11)

Fig. 2



Solar time dial showing the estimated 24 hr component of the daily variation due to atmospheric temperature in the ion-chamber data from Huancayo. Vectors 1 to 6 (as in table 1) and the overall mean *M* were found by the null method. The vectors (54-54) and *A* were derived by alternative methods (*vide text*). The circles denote standard deviations of the estimates.

for the years 1953 and 1954 satisfy this criterion and these data yield values $r=0.96$ and $\theta=0.94$, consistent with our earlier assumption $r, \theta \simeq 1$. Using these values of r and θ , the temperature wave in the Huancayo ion-chamber is found to have harmonic coefficients $a_1=0.026\%$ and $b_1=0.102\%$.

A further estimate of the temperature effect can be obtained by assuming that the primary variations produce 24 hr waves of identical magnitude and phase in the neutron monitor and the ionization chamber, or $r=\theta=1$. Here, the vector difference between the ionization chamber and neutron monitor data gives the temperature effect and the measurements made over the years 1953 to 1955 give values $a_1=0.035\%$ and $b_1=0.106\%$ for this vector. The estimates given by the last two methods are also shown in fig. 2 labelled (53-54) and A respectively. Although there is substantial agreement between the estimates given by the three methods, probably the most reliable one is obtained from the null method described first.

§ 3. METEOROLOGICAL DATA

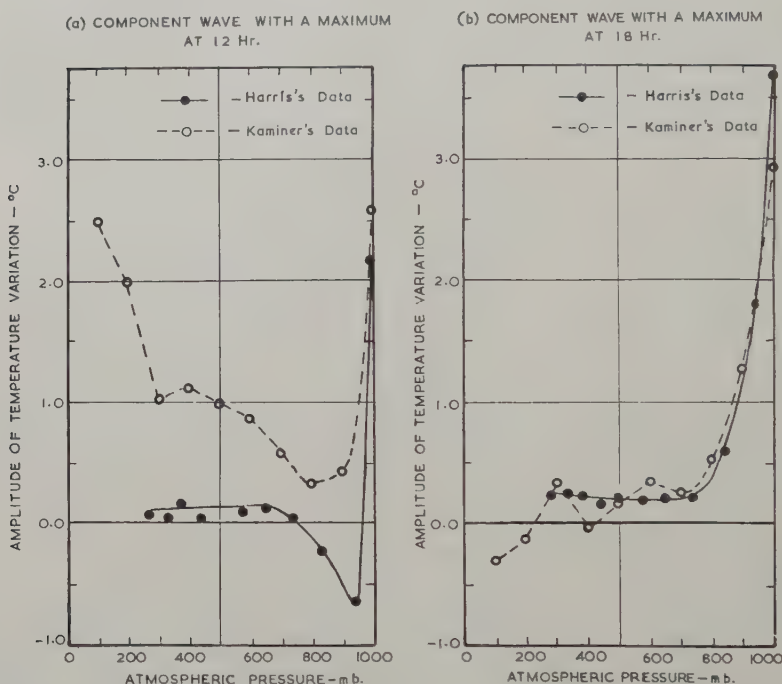
We shall now examine the evidence for a daily temperature variation in the atmosphere. The difficulties of measuring the true air temperature during daylight hours have already been indicated but in recent years temperature sensing devices with a special coating of very low absorptive power for the solar radiation spectrum have been developed (Brasefield 1948) and adopted for general use in routine upper air soundings in the U.S.A. An analysis of the upper air data obtained during the summer months of 1956 and 1957 at Washington, D.C. has been made by Harris (1959) who gives the diurnal and semi-diurnal components of the temperature variation in the atmosphere at altitude intervals of 1 km between 0.5 km and 9.5 km inclusive.

The atmospheric contribution to the solar daily variation of cosmic ray meson intensity was calculated by Dorman (1957) from the upper air data compiled by Kaminer and it is therefore interesting to compare these data from the U.S.S.R. with the more recent results of Harris. The data for the same season as Harris's results were analysed and differences between the two sets of results were found to exist in both amplitude and phase. The differences can best be seen by resolving each 24 hr wave of temperature into two components of the same periodicity, of which one has its time of maximum fixed at mid-day and the other one at 18 hr. Only the amplitudes of these components need be compared as the phases are constant. The advantage of this resolution is that the amplitude of only one of the components, viz. the one with a maximum at noon, will be sensitive to radiation errors in the measurements whilst the other will not be affected to any appreciable extent. This is because the radiation errors at times placed symmetrically with respect to noon should be roughly the same.

We have calculated the amplitudes of these components at various isobaric levels from the results of Harris and Kaminer and plotted them

in fig. 3 as functions of the atmospheric pressure. It can be seen from the diagram that the two sets of results give roughly the same amplitudes at various pressure levels for the component with an 18 hr maximum. Marked discrepancies are evident, however, in the other component, with the results of Harris showing a roughly constant amplitude of $\sim 0.1^\circ\text{C}$ between 750 mb and 350 mb while the Russian data give systematically larger amplitudes which increase with altitude from 0.5°C to about 1.2°C over the same range of the atmosphere.

Fig. 3

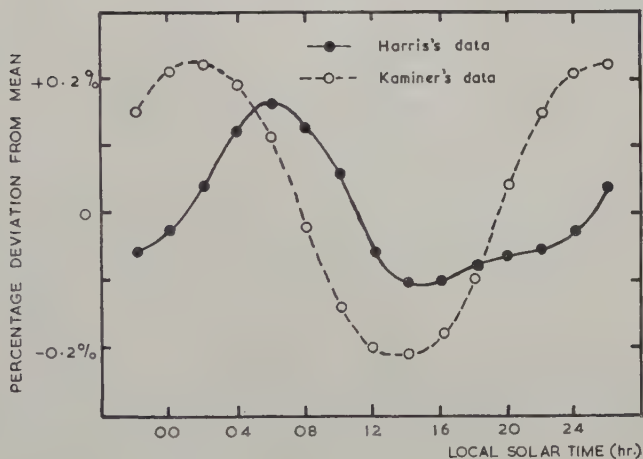


The diurnal temperature variation at various isobaric surfaces in the atmosphere according to the data of Harris (solid curve) and Kaminer (dotted curve). The amplitudes of the resolved components with a diurnal maximum at (a) 12 noon and (b) 18 hr L.T. are shown.

In determining the reasons for this divergence between the two results, factors such as the wide differences in the epochs of measurement and the geographical locations of the observing stations have to be taken into account. But there can be little doubt that at least a part of it is due to the differences in the relative amounts of radiation errors in the temperature measurements. The rough agreement between the two sets of data regarding the component which is not affected by radiation errors and the growth of the difference in the other component with increasing altitude appear to support this view.

We shall next examine the effects of the differences in aerological data on the estimates of the atmospheric contribution to the meson intensity. The temperature data of Harris together with the densities of temperature coefficients calculated by Dorman have been used to compute the percentage changes in the counting rate of a cubical meson telescope at sea-level originating in the portion of the atmosphere between sea-level and 250 mb, the limits being set by the availability of temperature data. The calculations give a first harmonic with an amplitude of 0.12% and a time of maximum at 05–30 hr, and a second harmonic with corresponding values of 0.04% and 06–30 hr respectively. The daily variation curve of meson intensity, which includes the effects of both harmonics, is shown in fig. 4 and a similar curve obtained by Dorman is also included for comparison purposes. The latter curve refers to the same season and the same range of the atmosphere, but is mainly diurnal with an amplitude of 0.22% and a maximum at 02 hr. As indicated above, the larger amplitude and the advance in phase from 05–30 hr to 02 hr must partly be due to the larger radiation errors in Kaminer's results.

Fig. 4



The solar daily variation due to atmospheric temperature in the counting rate of a cubical meson telescope at sea-level, calculated from the temperature data of Harris (solid curve). An earlier estimate of Dorman (dotted curve) is shown for comparison.

The contribution from that portion of the atmosphere above 250 mb cannot be estimated with any comparable accuracy but Kay (1951) has suggested that the temperature variation in the 150–80 mb region has an amplitude of $< 0.5^{\circ}\text{C}$, reaching a maximum value in the late afternoon, and if this is accepted as correct, our diurnal amplitude of 0.12% should be increased by about 0.03 to 0.05%.

As a result of the rapid attenuation of the diurnal amplitude with height evident in Harris's data, about two-thirds of the 0.12% amplitude originates in the atmosphere below the 750 mb isobar. It is well known, however, that the daily temperature variations in the lower troposphere show a marked seasonal dependence and vary considerably with geographical location. It can be expected, therefore, that the meson intensity variations will exhibit correspondingly large fluctuations from season to season at temperate latitudes and that the annual mean of the variation will have a smaller amplitude than the summer value of about 0.16% in agreement with the suggestion of Baliga and Thambyahpillai (1959). Moreover, the magnitudes of the temperature effects may vary substantially from one cosmic ray station to another.

We conclude, however, that the temperature wave deduced in § 2 from cosmic ray data is not inconsistent with that given by aerological measurements. Of course, Huancayo is at an altitude of 3350 m above sea-level and because of the lower mean energy of the μ -mesons recorded there, the densities of temperature coefficients should be somewhat larger than at sea-level, but this cannot lead to more than a fractional change in the amplitude of the expected variation. This should, however, be offset by the integration over a smaller range of the atmosphere (0 mb to \sim 650 mb) and the smaller amplitudes of the diurnal temperature variation in this range. These factors might explain the somewhat smaller amplitude of 0.11% deduced for the mountain station from cosmic ray measurements as compared with an amplitude of about 0.16% at sea-level. Moreover, the time of maximum of 05–40 hr obtained in § 2 appears to be borne out by the temperature data of Harris. The remarkable agreement between the two independent derivations with regard to the phase leaves little room for doubt that the diurnal maximum of the meson intensity variations due to atmospheric temperature occurs around 05–30 hr and not in the early hours of the morning, as has hitherto been believed (Dorman 1957).

§ 4. DIRECTIONAL MEASUREMENTS MADE AT MAWSON

In this section we shall consider the effects of correcting solar daily variation results for atmospheric influences using the temperature vector derived above. For this purpose we have selected the measurements made during the period May 1955 to February 1956 at Mawson, Antarctica ($67^{\circ}36'S$, $62^{\circ}53'E$ geographic) with a vertical telescope and another inclined at 45° to the vertical and rotated at regular intervals so as to look, in turn, due geographic north, east, south and west (Parsons 1957). The data, as published, were corrected only for changes of barometric pressure. These results were believed to be inconsistent with the view that the daily variation is the result of an anisotropic distribution of the primary flux, existing outside the earth's magnetic field (Parsons) but this conclusion should be re-examined after taking into account the contribution from atmospheric temperature changes. Moreover, the dipole approximation to the earth's magnetic field used by Parsons in his

discussion may not be sufficiently accurate for the purpose. Quenby and Webber (1959) have already pointed out the need for including the non-dipole part of the field in the calculation of cosmic ray cut-off rigidities and these terms are also of importance in calculating the asymptotic directions of all low energy cosmic rays.

The diurnal components of the solar daily variations measured by Parsons after the application of pressure corrections are shown in table 2 which gives the amplitudes and times of maxima for the various directions in which the telescopes were pointed. The features of interest in the results are the small amplitude of $0.014 \pm 0.016\%$ for the south direction as compared with amplitudes lying between 0.06 and 0.08% for the other directions and the appreciable spread in the times of maxima.

Table 2. The solar diurnal variations of cosmic-ray intensity at Mawson measured by telescopes pointing (a) in the vertical direction and (b) at zenith angles of 45° in various directions. The results corrected for barometer changes are shown together with the values obtained when a temperature corrections (vide text) is also included

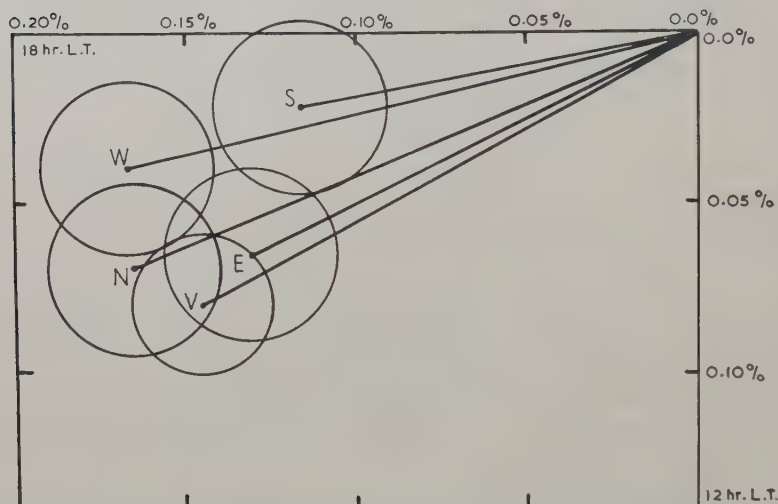
Telescope direction	Pressure corrected		Pressure and temperature corrected	
	Ampl. (%)	Time of max. (hr (L.T.))	Ampl. (%)	Time of max. (L.T.)
North	0.081 ± 0.016	14.50 ± 00.45	0.179 ± 0.025	16.28 ± 00.32
East	0.060 ± 0.016	13.20 ± 01.00	0.147 ± 0.025	16.12 ± 00.40
South	0.014 ± 0.016	13.50 ± 04.20	0.118 ± 0.025	17.12 ± 00.45
West	0.064 ± 0.016	16.10 ± 00.55	0.172 ± 0.025	17.08 ± 00.36
Vertical	0.079 ± 0.006	13.45 ± 00.15	0.166 ± 0.020	16.04 ± 00.30

In the absence of reliable information about the temperature changes over Mawson and the differences in temperature effects for vertical and inclined telescopes, we have used the temperature vector derived for the Huancayo ion chamber (amplitude 0.11%: phase 05-40 hr) to correct all five of the 24 hr waves obtained by Mawson. This procedure can give only a rough approximation to the actual state of affairs; but the uncertainty in the amplitude of the temperature correction has no marked effect on the conclusions to be drawn from the results. The diurnal variations obtained after the temperature correction are also given in table 2 and are shown on the harmonic dial in fig. 5. As a result of the correction, the ratio of the amplitudes in the south and north directions has been increased to about two-thirds and the spread in the phases for the various directions has been reduced.

In discussing his results in terms of a primary anisotropy, Parsons has adopted the suggestion of Brunberg and Dattner (1954) that the deflection

of the primary particles in the earth's magnetic field can be adequately taken into account by assuming a suitable mean value, between 20 GV and 40 GV, for the magnetic rigidity of the primary particles which are responsible for the daily variations. It is then possible to determine the asymptotic direction from which particles of this rigidity entered the earth's field before being deflected so as to be recorded by the appropriate counter telescope. This asymptotic direction is characterized by a latitude angle, measured from the geographic equatorial plane, and a longitude angle measured eastwards from the geographic meridian plane through Mawson. Using the data of Brunberg and Dattner (1953), based on the dipole approximation to the geomagnetic field, Parsons calculated these angles defining the asymptotic directions for the five telescope directions

Fig. 5

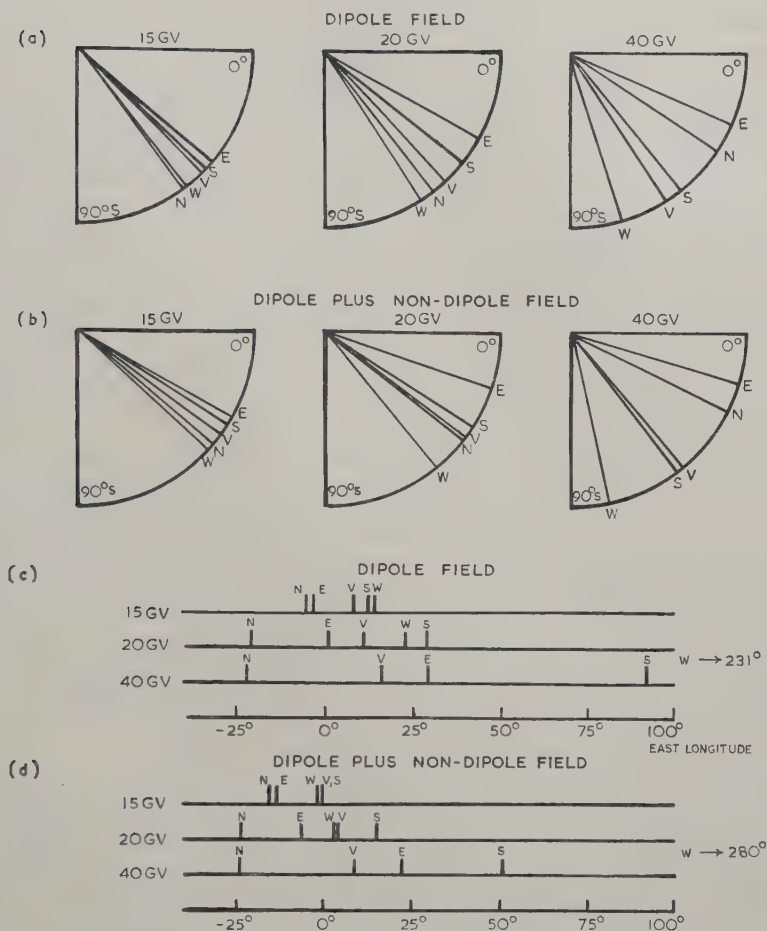


Harmonic dial representation of the cosmic ray diurnal variations at Mawson in the vertical direction (*V*) and at zenith angles of 45° in the north (*N*), east (*E*), south (*S*) and west (*W*) directions after correction for the effects of atmospheric pressure and temperature. Standard deviation circles are also shown.

used at Mawson and for particle rigidities of 15, 20 and 40 GV. The calculated angles of latitude and longitude are shown diagrammatically in fig. 6. Parsons has pointed out the following difficulties in reconciling his results (uncorrected for temperature effects) with an anisotropic distribution of the primaries which exists outside the earth's magnetic field. (a) Although the asymptotic latitude for the south pointing telescope was intermediate between those for the other telescope directions, the south telescope measured a much smaller amplitude than the other telescopes. This result cannot be explained easily because the strength of

the anisotropy can be expected to show a smooth variation with latitude. It will be seen from table 2 that this difficulty is largely removed by the application of a temperature correction. (b) The longitude angles are

Fig. 6



The angles of latitude and longitude defining the asymptotic directions from which primary particles enter the earth's magnetic field before being deflected so as to be recorded by the various telescopes at Mawson. Values for primary rigidities of 15, 20 and 40 GV are given.

- (a) and (b), asymptotic geographic latitudes;
- (c) and (d), asymptotic geographic longitudes (measured as in text);
- (a) and (c), take only the dipole part of the earth's magnetic field into account;
- (b) and (d), take into account the dipole and non-dipole parts of the earth's magnetic field.

different for the five telescopes and an excess of radiation entering the magnetic field from a particular direction should give rise to maxima in the measured intensity at different times. The sequence of the times of maxima should be in the order of decreasing longitude angle but the arrangement of measured times of maxima disagreed with the expected sequence.

Now, the geomagnetic latitude of Mawson on the centred dipole approximation is 73° , though the dip latitude, as defined by the angle of dip of the actual field lines, is 54° . Besides, the declinations of the horizontal components of the actual field and the field due to the centred dipole differ by 26° . Thus, the non-dipole part of the earth's magnetic field is quite large at Mawson and has to be taken into account. To do this a new dipole was fitted by requiring that the field of the fitted dipole gave a best fit to the actual field strength along the radius vector through Mawson. The angles of latitude and longitude corrected for the non-dipole field effects are also shown in fig. 6. It will be seen that there is no significant change in the distribution of latitude angles for 15 GV and 20 GV particles. Since the measured amplitudes, after temperature correction, are approximately equal within the errors, there is still no major inconsistency with the idea of a primary anisotropy.

The effect of including the non-dipole part of the earth's magnetic field is to bring the asymptotic longitudes for 15 GV and 20 GV closer to one another and this gives better agreement with the measured diurnal variations because the phases, after temperature correction, lie within a range of 20° while the standard deviations are about $\pm 10^\circ$ on each measurement.

In order to determine the direction and the mean rigidity of the primary anisotropy, the angle θ of the direction of maximum primary intensity measured eastwards from the sun-earth line was calculated from the equation :

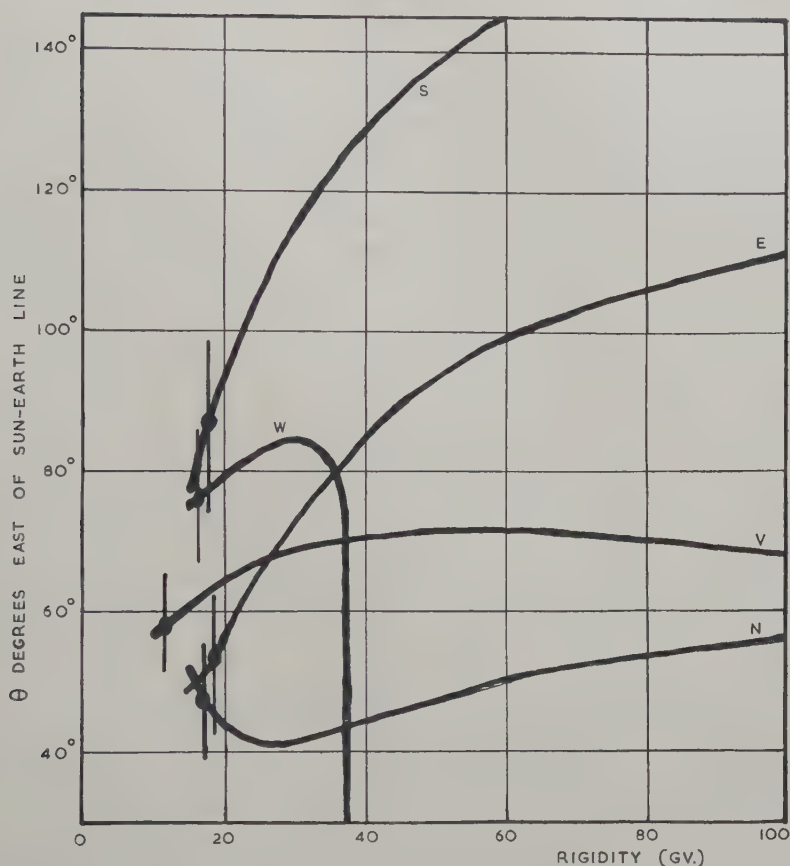
$$\theta = \psi_E + 15^\circ(t_{\max} - 12)$$

where ψ_E is the appropriate longitude angle corrected for non-dipole field effects and t_{\max} is the time of maximum in hours of the diurnal variation when all atmospheric influences have been eliminated. θ is plotted in fig. 7 as a function of the assumed mean rigidity of the primaries for the five telescope directions at Mawson. If the measured diurnal variations were due to a primary anisotropy, all curves should intersect at a point corresponding to the mean rigidity and the direction of the anisotropy. It will be seen from the diagram that all curves tend to converge at a primary rigidity of 10 GV and the curves separate out at higher rigidities. Thus, within the limits of statistical accuracy of the measurements, the results appear to be consistent with a primary anisotropy located roughly 70° due east of the sun-earth line, when the mean rigidity is taken to be about 10–20 GV. It is concluded, therefore, that the Mawson results can be reconciled with the idea of a primary anisotropy lying outside the earth's magnetic field, when the atmospheric influences and the deflecting action of the non-dipole part of the geomagnetic field are taken into account.

§ 5. CONCLUSIONS

Reasons have been given for believing that the corrections for atmospheric temperature effects on daily variation measurements of cosmic-ray intensity have not been evaluated accurately to date. Using purely cosmic-ray measurements, the atmospheric part of the solar diurnal variation has been found for the ionization chamber data from Huancayo.

Fig. 7



The longitude angle θ between the sun-earth line and the direction of maximum primary intensity outside the geomagnetic field, as given by each telescope at Mawson, plotted against the assumed values for the mean rigidity of the primaries producing the diurnal variation. The standard deviations of θ , which are independent of rigidity, are also shown.

This has an amplitude of 0.11% and a time of maximum around 05–40 hr L.T. Also, using some recent meteorological data, considered to be largely free from insolation errors, the atmospheric contribution has been calculated on the Dorman scheme for a cubical meson telescope operated at sea-level in temperate latitudes. The diurnal component here has an amplitude of

about 0.16% and a time of maximum at 05–30 hr during the summer season. Whilst the small difference in the amplitudes given by these two independent derivations is to be expected, the remarkable agreement in phase shows that the diurnal maximum of temperature in the troposphere is reached near 18 hr and not in the early afternoon as indicated by earlier radio-sonde measurements. When the latter calculation is compared with an estimate of Dorman for an ionization chamber at sea-level, it appears that the increased amplitude of 0.27% and the advance of phase to 02 hr found by Dorman are probably due to larger insolation errors in the aerological data used.

As an application of the calculated temperature effects, the directional telescope measurements made at Mawson were corrected using the Huan-cayo temperature vector. These corrected results are consistent with the view that the daily variation arises from a primary anisotropy outside the earth's magnetic field when the deflecting action of the geomagnetic field is taken into account. The results suggest that the primary radiation producing the daily variations possesses a mean rigidity in the range 10–20 GV and that the maximum intensity arrives from a direction 70° due east of the sun-earth line. A more comprehensive treatment of various aspects of the solar daily variation, taking into account the atmospheric effects discussed here, is now in preparation.

ACKNOWLEDGMENTS

We wish to thank Dr. H. Elliot for many valuable discussions and Professor P. Morrison for helpful comments. We are indebted to Mr. G. Wenk for the analysis summarized in fig. 1 and for calculations on the deflecting effects of the non-dipole part of the earth's magnetic field. The latter work by J. J. Quenby and G. Wenk will be published elsewhere.

REFERENCES

- BALIGA, S. P., and THAMBYAHPIILLAI, T., 1959, *Phil. Mag.*, **4**, 973.
 BRASEFIELD, H., 1948, *J. Meteorology, Washington*, **5**, 147.
 BRUNBERG, E. A., and DATNER, A., 1953, *Tellus*, **5**, Nos. 2, 135 and 3, 269;
 1954, *Ibid.*, **6**, No. 1, 73.
 CONFORTO, A. M., and SIMPSON, J. A., 1957, *Nuovo Cim.*, **5**, 1052.
 DORMAN, L. I., 1957, *Cosmic Ray Variations* (Moscow: State Publishing House).
 ELLIOT, H., and DOLBEAR, D. W. N., 1951, *J. atmos. terr. Phys.*, **1**, 205.
 HARRIS, M. F., 1959, *J. geophys. Res.*, **64**, 983.
 KAY, R. H., 1951, *Quart. J. R. met. Soc.*, **77**, 427.
 MEADA, K., and WADA, M., 1954, *J. sci. Res. Inst., Tokyo*, **48**, 71.
 MALMFORS, K. G., 1949, *Tellus*, **1**, 2.
 OLBERT, S., 1953, *Phys. Rev.*, **92**, 454.
 PARSONS, N. R., 1957, *Aust. J. Phys.*, **10**, 462.
 QUENBY, J. J., and WEBBER, W. R., 1959, *Phil. Mag.*, **4**, 90.
 TREFALL, H., 1955, *Proc. phys. Soc. Lond.*, **68**, 625, 893.
 VENKATESAN, D., and DATNER, A., 1959, *Tellus*, **11**, 116.
 WEBBER, W. R., and QUENBY, J. J., 1959, *Phil. Mag.*, **4**, 654.

Cosmic-ray Intensity Variations and the Interplanetary Magnetic Field†

By H. ELLIOT

Imperial College of Science and Technology, London

[Received January 25, 1960]

ABSTRACT

It is shown that the general characteristics of the cosmic-ray intensity variations can be explained in terms of a large-scale interplanetary magnetic field of predominantly dipole character but containing small-scale irregularities which act as scattering centres. It is suggested that the cosmic-ray data can be taken as evidence for the existence of such a field. The strength of the field, which is generated by current systems in the solar corona, is dependent on the level of solar activity but must in general be in the region of 10^{-5} to 10^{-4} gauss at the earth's orbit.

§ 1. INTRODUCTION

THE primary cosmic rays show certain characteristic variations in intensity which are rather closely correlated with solar and geomagnetic activity and it is generally believed that this relationship exists because of the modulation of the cosmic-ray flux by changing magnetic, and possibly electric, fields in interplanetary space. The central problem in the study of cosmic-ray intensity variations is that of constructing a detailed picture of these fields from the experimental data available and it is the purpose of this paper to discuss one particular model which appears to fit these data rather well. For any model to be acceptable, the configuration of fields chosen must be capable of explaining the eleven-year variation in cosmic-ray intensity over the sunspot cycle, the Forbush type decrease usually associated with great magnetic storms, the quasi-periodic 27-day variation, and the solar daily variation. Finally, it must be compatible with the propagation characteristics from sun to earth of particles accelerated to cosmic-ray energies in solar flares.

During recent years several models of the interplanetary electric and magnetic fields have been proposed but none has so far proved completely satisfactory. Broadly speaking, these proposals may be divided into two categories; those which depend on randomly orientated fields of a kind generated by the turbulent motion of highly ionized gas (Morrison 1955, Meyer *et al.* 1956, Singer 1957, Parker 1958) and those which depend on rather highly organized and uniform magnetic fields associated with solar streams (e.g. Alfvén 1954, Brunberg and Dattner 1954, Cocconi *et al.* 1957, Dorman 1957). Since these magnetic fields vary with time it follows that there must also be electric fields associated with them and in some models these electric

† Communicated by the Author.

fields play an important, even dominant, role. The possibility of an extensive interplanetary electric field capable of modulating the cosmic ray intensity has recently been discussed by Ehmert (1959).

The modulating mechanism which will be discussed here has some features in common with these earlier models but differs from them in that it depends primarily on the existence of a general solar magnetic field which extends far out into interplanetary space beyond the earth's orbit and which is essentially non-turbulent. The screening effect of a solar dipole field on the cosmic-ray intensity at the earth was first discussed by Jánossy (1937) many years ago when it was believed that the sun possessed a dipole moment of $\sim 10^{34}$ gauss cm³, corresponding to a polar field of about 30 gauss. We now know that the solar dipole field, as measured in the photosphere, is much smaller than this (~ 1 gauss at the pole) and that it is variable (Babcock 1959). Such a field would not in itself be capable of producing any appreciable effect on the cosmic-ray intensity at the earth. If, however, large-scale current systems exist in the solar corona the effective dipole moment of the sun may be much greater than would appear from measurements of Zeeman splitting of spectral lines originating in the photosphere. In the present model it is suggested that such currents do exist and that the field they produce is responsible for the cosmic-ray intensity variations. If we are to make a quantitative estimate of the effects of such a field on the cosmic rays, it is necessary to idealize and a close relation between the model and reality must not be expected at this stage.

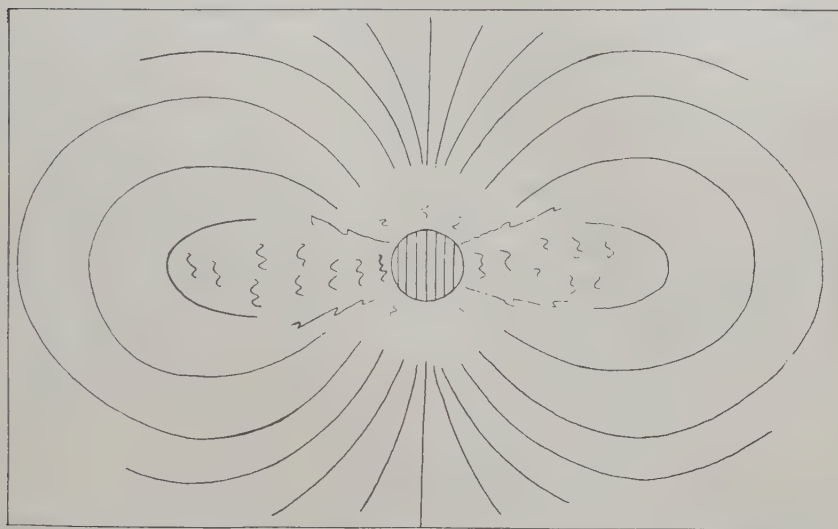
§ 2. PROPOSED MODEL OF THE SOLAR MAGNETIC FIELD

In attempting to build up a picture of the interplanetary field from the cosmic-ray evidence we must begin with the simplest configuration capable of accounting for the qualitative behaviour of the cosmic rays and then elaborate it to the extent merited by the observational detail available. There are two simple models which lend themselves readily to quantitative treatment. In the first of these it is supposed that the field changes rapidly in intensity and direction as a function of space coordinates. This corresponds physically to a highly turbulent field in which the motion of the cosmic-ray particles can be treated as a diffusion problem without the necessity of any knowledge about details of their trajectories. This model has already been extensively discussed in the references given above.

At the other extreme, we can suppose that the current systems generating the field lie predominantly within the earth's orbit and possess a high degree of axial symmetry about the sun generating a nearly dipole field at points outside the earth's orbit. The motion of cosmic-ray particles in a dipole field is well understood and a quantitative treatment of this model is also possible. Close to the sun the field must be very disturbed and far from regular because of patchy photospheric fields, sunspot fields and the violent hydromagnetic motions of the solar atmosphere close to the surface. The occurrence of magnetic storms can be taken as evidence that these violent disturbances sometimes extend out as far as the earth's orbit

and presumably beyond it. For the purpose of calculation we shall divide the solar field schematically into three regions as shown in fig. 1, an inner region in which the field is much disturbed and which is surrounded by another region in which the large scale axially symmetrical currents flow. These currents generate a dipole field which constitutes the third and outer region extending far beyond the earth's orbit. This dipole field will be periodically disturbed by the outward moving solar streams responsible for magnetic disturbances and auroral activity.

Fig. 1



Schematic representation of the magnetic field generated by currents flowing in the solar corona.

§ 3. EFFECT OF THE SOLAR FIELD ON THE COSMIC-RAY INTENSITY

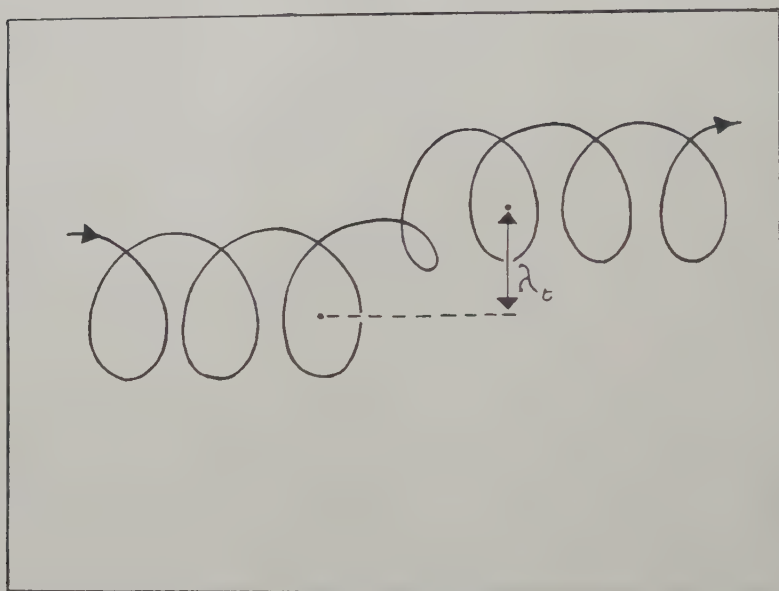
If the field outside the earth's orbit was a completely unperturbed dipole field we should have the familiar allowed and forbidden regions given by the Störmer theory. That is to say, we should have some minimum magnetic rigidity, p_{\min} below which no particle could reach the earth and a maximum p_{\max} above which particles could arrive from all directions. In the intervening range between p_{\min} and p_{\max} the arrival of particles would be restricted to certain directions. The values of p_{\max} and p_{\min} are given by

$$p_{\max} = \frac{300 M_s}{Z r_e^2} \quad \text{and} \quad p_{\min} = (\sqrt{2} - 1)^2 \frac{300 M_s}{Z r_e^2} \text{ in gv} \quad . \quad . \quad (1)$$

where r_e is the radius of the earth's orbit in centimetres (assumed circular), M_s is the dipole moment of the solar current system in gauss cm³, Z is the particle charge in units of the electron charge e . For a singly charged particle $Z = 1$ and the magnetic rigidity in gv is numerically equal to the particle momentum in GeV/c.

In reality the situation is more complicated than this because of the perturbations introduced by outward-moving solar streams. If, as seems likely, these streams can produce gradients in the field of a sufficient magnitude, particles will be scattered into forbidden regions which would otherwise be empty according to the Stormer theory. The requirement for such scattering to be effective is that the field should change by a substantial fraction in a distance corresponding to the radius of curvature of the particle trajectory in the unperturbed field. Alternatively the perturbing field, being time dependent, should change appreciably during the time the particle requires to cross the perturbing region.

Fig. 2



Scattering of cosmic-ray particle by irregularity in dipole field.

§ 4. SCATTERING BY MAGNETIC IRREGULARITIES

Near the equatorial plane of the solar dipole, where scattering by solar streams is most likely to occur, a cosmic-ray particle describes a nearly helical path with its axis perpendicular to the ecliptic plane. If such a particle encounters a sharp gradient in the field the radius of curvature of its trajectory at that point is changed and the particle is scattered into another orbit. As a result, the guiding centre is scattered through some distance λ_i . The projection of the trajectory in the equatorial plane is shown in fig. 2.

A detailed discussion of the scattering process is only possible if we have detailed information about the structure of the scattering region. In the absence of such information we can only make a rough schematic treatment as follows.

We suppose that the field strength in the irregularity is of the same order as that outside and that the physical dimensions of the irregularity are small compared with ρ , the radius of curvature in the unperturbed field. This means that on passing through the irregularity the particle suffers a small angular deflection

$$\theta \sim \frac{300 B_i l}{p}$$

where B_i is the mean field strength along the particle trajectory in the perturbing field and l is its length.

Then $\lambda_t \sim \theta \rho \sim (B_i/B) l$ where B is the local value of the unperturbed field. If we suppose that the perturbing field B_i is linearly related to B , so that $B_i/B = \text{constant}$, it follows that λ_t is independent of particle rigidity and of position. We can then regard the motion of the cosmic-ray particle in the solar dipole field as a combination of that described by Störmer theory and diffusion with a constant transport mean free path given by λ_t above.

In the absence of absorption, the effect of the diffusion would be to fill all the forbidden regions with particles which, by Liouville's theorem, would eventually attain the same intensity as that existing in interstellar space. In fact, as we shall see, absorption by the sun reduces the intensity in the forbidden regions below that existing outside the solar system and an equilibrium intensity is reached which is dependent on the relative rates of diffusion and absorption.

§ 5. ABSORPTION OF COSMIC-RAY PARTICLES BY THE SUN

In order to estimate the extent to which particles are absorbed by the sun during their diffusion inwards we have to take into account the complete field distribution and not merely the dipole part. The simplest assumption we can make about the inner part of the field is to suppose that it is generated by a spherical current shell of radius r_c in which the currents flow along lines of latitude. If the current strength is proportional to $\cos \lambda$, where λ is the latitude, such a current system generates a dipole field outside the shell and a uniform field within it as in fig. 3.

At each scattering collision there is a finite probability that the diffusing particle can be scattered into such a direction that it will travel directly along the field lines into the sun and be absorbed. Thus, in a sense, we may regard the solar absorber as spread throughout the solar system. If the particle is to reach the sun in this way it must be scattered into a small solid angle Ω_A as shown in fig. 3, such that $\Omega_A = 2\pi (1 - \cos \psi)$ where ψ is given by the 'trapping condition', $\sin^2 \psi = B/B_0$. That is to say, a particle with a pitch angle greater than $[(\pi/2) - \psi]$ in the equatorial plane is transmitted at the surface of the sphere radius r_c and may enter the turbulent region and finally strike the sun. A particle with a pitch angle less than $[(\pi/2) - \psi]$, on the other hand, will be reflected before reaching the surface at $r = r_c$.

We must now relate this absorption probability to a particular line of force in the external dipole field. The dipole line of force which connects to E cuts the equatorial plane at C at a distance

$$r = \frac{r_c}{\cos^2 \lambda} = \frac{r_c^3}{r'^2}$$

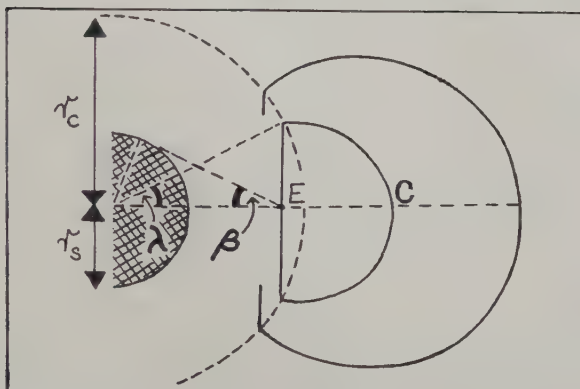
and therefore we can write

$$P_a(r) = P_c \frac{A_s r}{r_c^3}.$$

Combining this with the expression for Ω_A , we obtain for the probability of absorption per scattering collision

$$\left. \begin{aligned} P(r) &= \frac{2\Omega_A}{4\pi} P_a(r) = P_c \frac{A_s r}{r_c^3} \{1 - \sqrt{1 - B/B_0}\} \\ \text{i.e. } P(r) &\sim P_c \frac{A_s}{2r^2} \quad \text{on putting } B/B_0 = \left(\frac{r_c}{r}\right)^3 \end{aligned} \right\} \quad (4)$$

Fig. 4



The geometry of the inner field region close to the sun.

§ 6. QUANTITATIVE DISCUSSION OF THE SCREENING BY A SOLAR DIPOLE FIELD

Our problem, then, is to calculate the cosmic ray intensity inside a diffusing medium with a transport mean free path λ_t and a probability of absorption per collision given by P in equation (4). In order to do so we must first define the outer boundary to the diffusing region. If M_s is the magnetic moment of the solar current system and λ the latitude angle measured from the equatorial plane, then according to Störmer theory, on the surface of revolution defined by

$$r = \sqrt{\left(\frac{300 M_s}{p}\right) \cdot \frac{\cos^2 \lambda}{[1 + (1 - \cos^3 \lambda)^{1/2}]}} \quad (5)$$

particles can arrive from all directions without restriction and on this surface cosmic rays of rigidity p or greater will have the full galactic intensity.

At first sight it might seem that this surface should be taken as the outer boundary of the diffusing region. In fact it is much more realistic to take instead the surface on which particles cannot arrive directly from any direction. Inside this surface particles can only arrive by diffusion whereas, as we shall see in § 9, the flux outside this surface is very closely isotropic and of nearly full galactic intensity.

This surface, S_0 , for which direct access is forbidden, is the surface of revolution generated by the curve.

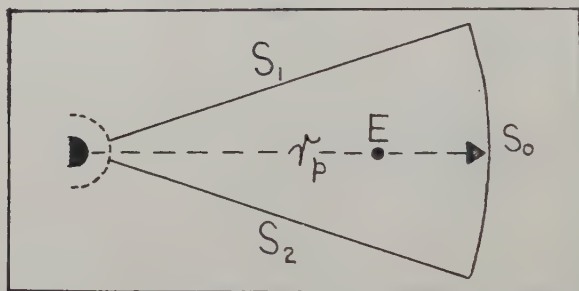
$$r = \sqrt{\left(\frac{300 M_s}{p}\right) \cdot \frac{\cos^2 \lambda}{[1 + (1 + \cos^3 \lambda)^{1/2}]}}.$$

For λ between $\pm 30^\circ$ we can take this to be part of the surface of a sphere of radius r_p where r_p is the value of r corresponding to $\lambda = 15^\circ$. Then

$$r_p = \sqrt{\left(\frac{45 M_s}{p}\right)} \quad \dots \dots \dots (6)$$

and the diffusing region in the dipole field is bounded by the surfaces S_0, S_1, S_2 (fig. 5).

Fig. 5



The boundaries of the region in which diffusion occurs.

Provided the absorption probability is small the net flux crossing the surfaces S_1 and S_2 , fig. 5, is very nearly zero and consequently we can regard the system as spherically symmetrical so far as E is concerned.

The problem is therefore simplified to that of calculating the intensity distribution inside a diffusing spherical medium of radius $r_p = \sqrt{(45 M_s/p)}$ with a transport mean free path $\lambda_t = B_t l/B$ and a probability of absorption per collision $P = P_c (A_s/2r^2)$. If N is the number of scattering collisions per second and n the number of cosmic rays per unit volume, the appropriate diffusion equation is

$$N(\lambda_t^2/3) \nabla^2 n(r, p, t) - NP(r) n(r, p, t) = \frac{\partial n(r, p, t)}{\partial t} \quad \dots \dots (7)$$

We are interested in the stationary solution and we therefore set $\frac{\partial n}{\partial t} = 0$ and put ϕ , the differential flux of particles of rigidity p , equal to nv . Then we have

$$(\lambda_t^2/3) \nabla^2 \phi(r, p) - P(r) \phi(r, p) = 0 \quad \dots \dots \dots (8)$$

or
$$\nabla^2 \phi(r, p) - K \frac{\phi(r, p)}{r^2} = 0 \quad \text{where} \quad K = \frac{3P_c A_s}{2\lambda_t^2}.$$

The solution of this equation is

$$\phi = Ar^{-\frac{1}{2}\{1-\sqrt{(1+4K)}\}} + Br^{-\frac{1}{2}\{1+\sqrt{(1+4K)}\}}$$

but since ϕ must be finite at $r=0$ it follows that $B=0$. Finally using the other boundary condition, $\phi=\phi_\infty$ at $r=r_p$, we obtain

$$\begin{aligned}\phi &= \phi_\infty \left[\frac{r_p}{r} \right]^{\frac{1}{2}\{1-\sqrt{(1+4K)}\}} \\ &= \phi_\infty \left[\frac{\sqrt{(45M_s)}}{r_p^{1/2}} \right]^{\frac{1}{2}\{1-\sqrt{(1+4K)}\}} \quad \dots \quad (9)\end{aligned}$$

The intensity at the earth is therefore given by

$$\phi_E = \phi_\infty \left[\frac{\sqrt{(45M_s)}}{r_e p^{1/2}} \right]^{\frac{1}{2}\{1-\sqrt{(1+4K)}\}} \quad \dots \quad (10)$$

where $r_e = 1.5 \times 10^{13}$ cm and $r_e^2 p \leq 45 M_s$.

In the calculation so far, we have taken λ_t to be constant but this cannot be so at low values of p when the scattering angle, θ approaches $\pi/2$. Under these circumstances, that is to say for large θ , $\lambda_t \rightarrow \rho$. We can take this into account approximately by writing

$$\lambda'_t = \lambda_t \frac{p}{p_0 + p} \quad \text{where} \quad \lambda_t = \frac{B_t l}{B} \quad \text{and} \quad p_0 = 300 B_t l.$$

This gives λ'_t constant and equal to $B_t l/B$ at large values of p and $\lambda'_t = \rho$ at $p \ll p_0$. Consequently in (10) we can replace K by

$$K' = \frac{3P_c A_s (p_0 + p)^2}{2 \lambda_t^2 p^2} = K \left[\frac{p_0 + p}{p} \right]^2$$

giving

$$\phi_E = \phi_\infty \left[\frac{\sqrt{(45M_s)}}{r_e p^{1/2}} \right]^{\frac{1}{2}\{1-\sqrt{(1+4K[(p_0+p)/p]^2})\}} \quad \dots \quad (11)$$

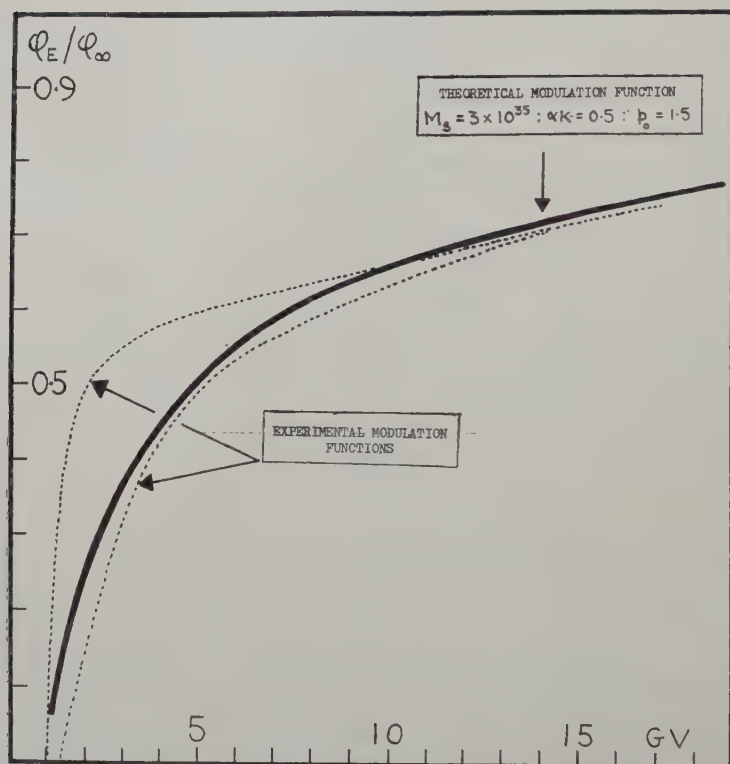
§ 7. THE 11-YEAR VARIATION IN INTENSITY

Equation (11) gives the steady-state cosmic-ray flux at the earth in terms of the galactic flux ϕ_∞ , as a function of magnetic rigidity p . It seems likely that M_s will vary over the solar cycle and so will the probability of capture by the sun, P , because of its dependence on Ω_A (eqn. (3)). In deducing the value of Ω_A used above we have assumed a strictly dipole field and therefore a particular relationship between B and B_0 (eqn. (2)). At the time of a major solar disturbance, however, this field must be seriously distorted by current systems associated with the outward moving solar material, indeed it has been suggested by Cocconi *et al.* (1957) that at such a time there may be a radial magnetic field connecting the earth and sun. In this extreme case, or for any increase in field strength in the region of solar emission between $\lambda = \pm 30^\circ$, the ratio B/B_0 is increased and so is the value of Ω_A . It follows from (3) that P is also increased and consequently at sunspot maximum, when the frequency of large disturbances is greatest, the average value of K must be greater than that appropriate to a dipole field becoming say αK , where α is a numerical factor greater than 1. According

to the present model it is this increase in both K and M_s with increasing solar activity which is responsible for the reduction in intensity of low-energy cosmic rays. The most convenient way of relating the predictions of the model to experimental observations is to compare the predicted variation of the ratio ϕ_E/ϕ_∞ with rigidity to that actually observed.

The differential rigidity spectrum at the earth, $\phi_E(p)$, can be measured at any particular time during the solar cycle but since we do not know $\phi_\infty(p)$, the galactic spectrum, we cannot deduce the ratio $\phi_E(p)/\phi_\infty(p)$

Fig. 6

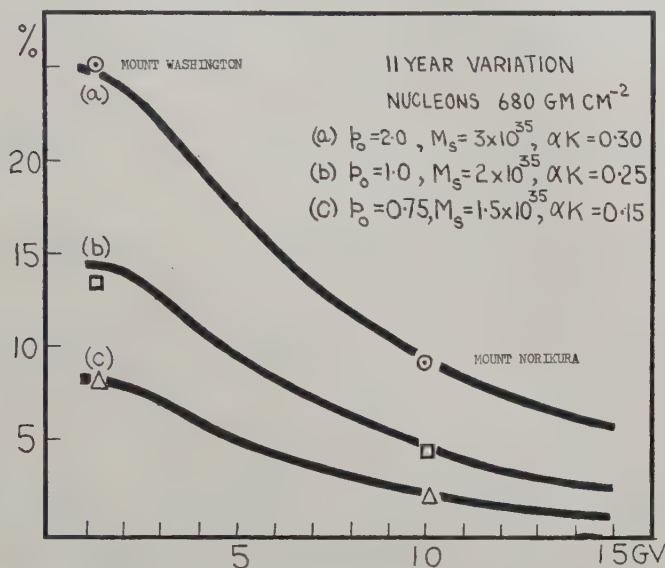


Comparison between calculated rigidity dependence of the 11-year variation and that deduced from observation.

directly from observation. There are however, two alternative plausible assumptions we can make which enable us to make a rough comparison between theory and observation. We can suppose that at sunspot minimum $K \sim 0$ in which case the galactic spectrum is simply the differential spectrum observed at the earth at that time. Alternatively, we can suppose that the galactic spectrum follows the same power law as that observed at the earth for high rigidity particles which are not affected by the solar modulation process.

The function $[1 - \phi_E/\phi_\infty]$ deduced from experimental observations for each of these two assumptions (Elliot *et al.* 1959) is shown in fig. 6 together with that calculated for $M_s = 3 \times 10^{35}$, $\alpha K = 0.5$ and $p_0 = 1.5$ GV. It will be seen that there is reasonable agreement between theory and observation for these values of the three parameters. Using these values of αK and M and p_0 we can, with the aid of the differential response functions obtained by Webber and Quenby (1959), deduce the ratio of the change in intensity over the solar cycle for the nucleonic component and for the μ -meson intensity recorded at sea level. The calculated ratio of 3 is in rough agreement with the observed value of ~ 4 at Ottawa (Fenton *et al.* 1958).

Fig. 7



Comparison between calculated and observed rigidity dependence of the 11-year variation for nucleons at mountain altitudes.

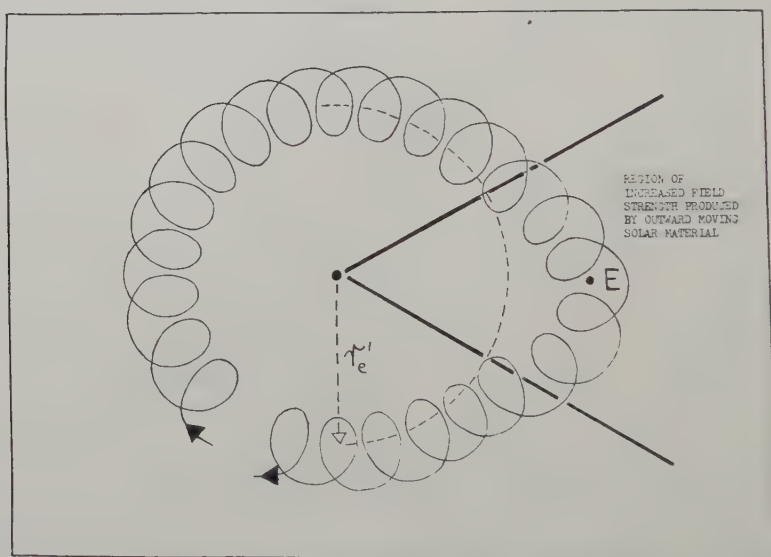
A more detailed comparison between theory and observation is shown in fig. 7, where the calculated dependence of the long-term intensity variation on station cut-off rigidity is shown together with the observed changes at mountain stations for three different epochs of the solar cycle. It will be seen that the theoretical curves fit the data quite well provided we assume a systematic decrease in M_s , αK and p_0 with decreasing solar activity. In this calculation it has been assumed that the galactic spectrum is the same as that observed at the earth at sunspot minimum.

§ 8. FORBUSH-TYPE INTENSITY DECREASES

At the time of a great magnetic storm it frequently happens that the cosmic-ray intensity at sea level drops sharply by from 5 to 10% and then slowly recovers, over a period of several days, back to the pre-storm value.

According to the present model these Forbush decreases are closely connected with the 11-year cycle in intensity and are observed when the earth lies directly in the path of the outward-streaming material emitted by an active region on the sun. If we suppose that this solar material to some extent carries the field with it then the trapped cosmic-ray particles spiralling round the field lines must be convected outwards at the same time. It follows that on such occasions an observer at the earth will sample a sub-normal cosmic-ray flux characteristic of that existing under normal conditions, at some point lying nearer to the sun. Alternatively, if we regard the outward motion of the solar gas as leading to an increased field strength in the region through which it passes illustrated in fig. 8, it follows that because $\text{grad } (B)$ now has a tangential component, the cosmic-ray particles trapped in the field will tend to drift round the region of higher field, producing the same result, a decrease in intensity.

Fig. 8

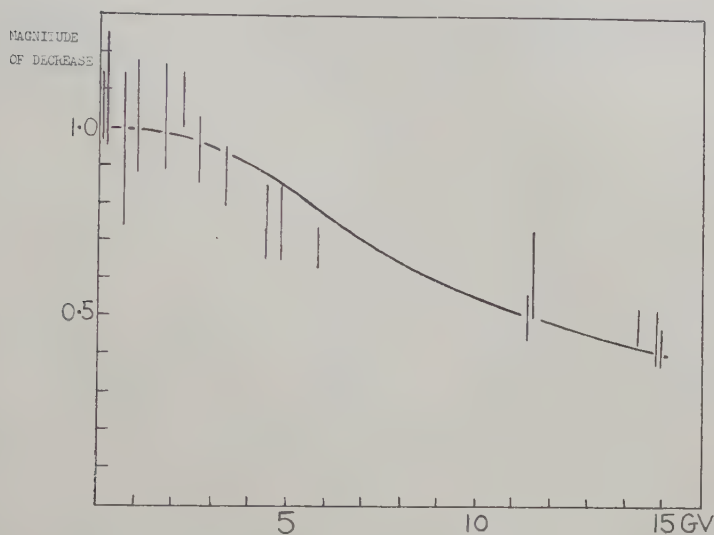


The outward convection of trapped cosmic radiation at the time of a Forbush decrease.

This will not be the only effect however. An increase in the field strength near the ecliptic plane will also lead to a greater probability of particles being absorbed by the sun. This arises from the increase in the ratio B/B_0 , and therefore in Ω_A , compared with that for the unperturbed dipole field. As pointed out in the previous section, the superposition of many such events will be contributory towards the reduction in cosmic-ray intensity at sun-spot maximum. Because of the time scales involved, however, this change in absorption probability is not very important in discussing the Forbush decrease and will be ignored in the following treatment.

We can obtain a rough idea of the form of the primary rigidity spectrum during a Forbush event in the following way. According to fig. 8, during the period of reduced intensity the cosmic rays received at the earth in the rigidity range $p \leq (45M_s/r_e^2)$ will be those normally trapped in orbits which do not extend out beyond some distance r_e' from the sun. This distance r_e' will be related to r_e in some rather complicated way, which we shall not attempt to deduce here, by the magnitude of the perturbing field.

Fig. 9



Comparison between observed and calculated change in intensity of the nucleonic component at sea level during the Forbush decrease of March 25/26, 1958. The intensity changes have been normalized to the value 1.0 at high latitudes.

It is sufficient for the present, to point out that at the time of the Forbush decrease we should observe a spectrum of the form

$$\phi_E'(p) = \phi_\infty(p) \left[\frac{\sqrt{45 M_s}}{r_e' p^{1/2}} \right]^{\frac{1}{2} \{1 - \sqrt{1 + 4\alpha K[(p_0' + p)/p]^2}\}} \quad (12)$$

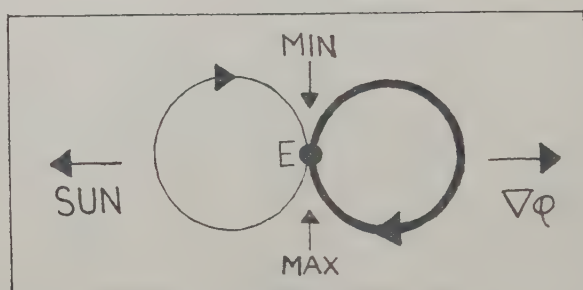
where r_e and p_0 have been replaced by r_e' and p_0' , and p_0' is greater than p_0 because of the greater value of the field at r_e' . In fact $p_0' = p_0 (r_e/r_e')^3$, and if we take $r_e/r_e' = 1.22$, a value which gives a decrease in intensity of 5.8% in the nucleonic component at sea level at high latitudes, we find $p_0' = 3.7$ gv. Using these values for r_e' and p_0' and taking $\alpha K = 0.3$ and $M_s = 10^{35}$, we can calculate $\phi_E'(p)$ and the function $[1 - \phi_E'(p)]$. Using this, together with the appropriate differential response functions, we can calculate the corresponding change in the μ -meson component which turns out to be 2.65%. The ratio of 2.2 between the changes in the nucleonic and meson components agrees well with the mean value of about 2.3

obtained by Fenton *et al.* (1958) for the Resolute and Ottawa stations in 1957. The change in intensity of the nucleonic component at sea level has also been calculated as a function of station magnetic cut-off rigidity using eqns. (12). The calculated dependence is shown in fig. 9 together with the observed latitude dependence for the Forbush decrease of March 25/26, 1958. Again the agreement is satisfactory.

§ 9. THE SOLAR DAILY VARIATION

In the discussion so far we have assumed the cosmic-ray flux at the earth to be isotropic at all rigidities, but in fact a large-scale solar dipole field of the kind discussed above leads to a small anisotropy which would appear as a solar daily variation in intensity. This anisotropy arises in the following way.

Fig. 10



Anisotropy due to the gradient in the cosmic-ray flux.

In a uniform magnetic field in which there exists a gradient in the particle number density the flux at a point E will be anisotropic as shown in fig. 10. The fractional magnitude of the anisotropy is given by $\delta = 2\rho (\text{grad } \phi)/\phi$ and an observer rotating with the earth at E and receiving radiation from directions close to the ecliptic plane would measure a sinusoidal variation in intensity with amplitude $\delta/2 = \rho (\text{grad } \phi)/\phi$.

From (12) we have

$$\delta/2 = - \frac{pr_e^2}{600 M_s} \left[1 - \sqrt{\left(1 + 4K\alpha \left\{ \frac{p_0 + p}{p} \right\}^2 \right)} \right].$$

Taking representative values for the three parameters $M_s = 2 \times 10^{35}$, $\alpha K = 0.25$ and $p_0 = 1$, we can calculate $\delta/2$ as a function of p and this is shown in fig. 11 for $1 < p < 20$ gv. This calculation is valid in so far as the motion of the cosmic ray particles can be treated as a diffusion problem but it certainly breaks down for $p > 30$ gv where ρ becomes comparable with the barrier thickness. For values of $p > 48$ gv, however, we can obtain a rough estimate of the degree of anisotropy in the following way.

If the solar field was strictly dipolar in character, particles with rigidity less than

$$p_{\min} = (\sqrt{2} - 1)^2 \frac{300 M_s}{Z r_e^2}$$

could only reach the earth as a result of scattering, whereas those with rigidity greater than

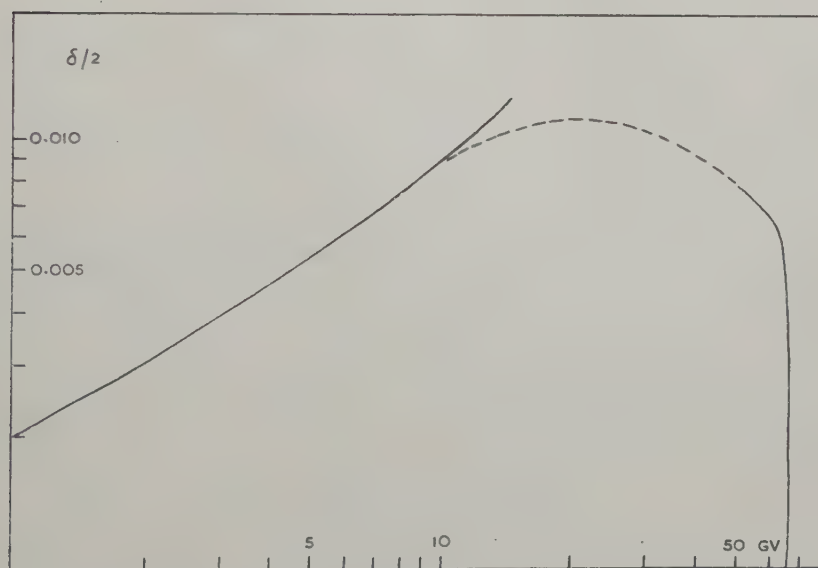
$$p_{\max} = \frac{300 M_s}{Z r_e^2}$$

could arrive freely from all directions. In the range between p_{\min} and p_{\max} singly charged particles could only arrive from directions lying outside a cone of semi-angle γ given by

$$\cos \gamma = 2 \sqrt{\left(\frac{300 M_s}{p r_e^2} \right) - \frac{300 M_s}{p r_e^2}} \quad . \quad . \quad . \quad . \quad (13)$$

and with its axis in the plane of the ecliptic and perpendicular to the sun-earth line (fig. 12).

Fig. 11

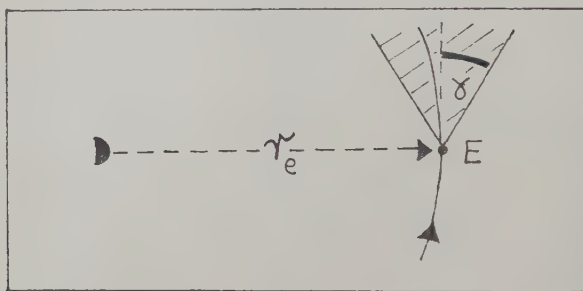


The calculated rigidity dependence of the solar daily variation.

For $M_s = 2.0 \times 10^{35}$, $p_{\min} = 4.6 \times 10^{10}$ and $p_{\max} = 2.7 \times 10^{11}$ so that for this value of M_s we might expect particles in this rigidity range to be anisotropically incident on the earth. Maximum contribution to the daily intensity variation would be expected from those particles for which γ lies between about 45° and 135° , that is to say from particles with $48 \text{ GV} < p < 112 \text{ GV}$. Taking the direction of the solar dipole to be in the same sense as that of the earth and neglecting the deflection in the earth's

magnetic field, maximum intensity would occur at 18 hours local time and minimum intensity at 06 hours. Primary particles in the rigidity range 48 to 112 gv produce about 20% of the μ -meson flux at sea level and if the solar field was entirely dipolar in character we should expect a large daily variation with an amplitude $\sim 10\%$. Because of scattering, however, the forbidden directions will to a large extent be filled with trapped particles and the magnitude of the anisotropy much reduced (Alfvén (1947), Kane *et al.* (1949), Treiman (1954)). The intensity in the forbidden directions will be determined by the probability of absorption by the sun and can be estimated very roughly as follows.

Fig. 12



The Stormer forbidden cone at the earth due to the dipole field.

In order that a particle in the momentum range we are considering shall be transferred from an allowed into a forbidden direction it has to be scattered through an angle comparable to γ given by (13). This angular deflection will be made up for a number n , of small deviations $\theta = 300 B_i l / p$ such that

$$\cos^{-1} \left\{ 2 \sqrt{\left(\frac{300 M_s}{p r_e^2} \right) - \frac{300 M_s}{p r_e^2}} \right\} = n^2 \frac{300 B_i l}{p}.$$

From eqn. (4) and § 7 the probability of absorption per collision, P is given by $P = P_c (A_s \alpha / 2 r_e^2)$, and taking

$$\alpha K = \frac{3 P_c A_s \alpha}{2 l^2} \left(\frac{B}{B_i} \right)^2 = 0.25,$$

as indicated by the 11-year variation, we find

$$P = \left(\frac{B_i}{B} \right)^2 \frac{l^2}{12 r_e^2}.$$

The intensity in the forbidden direction at the earth, ϕ_F , is therefore given by

$$\phi_F = \phi_\infty \left(1 - \frac{p^2}{12 (300 B r_e)^2} \left[\cos^{-1} \left\{ 2 \sqrt{\left(\frac{300 M_s}{p r_e^2} \right) - \frac{300 M_s}{p r_e^2}} \right\} \right]^2 \right). \quad (15)$$

Putting in the value of B at the earth's orbit appropriate to $M_s = 2 \times 10^{35}$ and taking $p = 5 \times 10^{10}$ for instance, we obtain $(\phi_\infty - \phi_F) / \phi_\infty = 0.016$,

indicating that for this rigidity the intensity in the forbidden directions is only 1.6% less than that in the allowed directions. At $p > 67$ gv the forbidden cone has a half angle, γ , less than $\pi/2$ and in consequence can be filled to the full galactic intensity because particles can be scattered into it with a negligible probability that they will strike the sun in the process. It follows that the magnitude of the anisotropy, $\delta/2$, in primary intensity must fall from $\sim 0.8\%$ at 50 gv to zero at $p = 67$ gv as shown in fig. 11.

In the intervening region between 20 gv and 48 gv the extent of the anisotropy is very uncertain but it seems not unreasonable to join the two branches of the curve smoothly together by the dashed line in fig. 11. Any attempt to calculate the magnitude of δ for this part of the spectrum requires a more detailed specification of the field than has been given so far and it seems likely that it is just this dependence on detail which is responsible for the great variability of the daily variation.

If, however, for the present, we accept the curve in fig. 11 as giving an approximate estimate of the degree of anisotropy for particle trajectories which lie in or near the plane of the ecliptic, we can use it to calculate the expected diurnal variation in intensity of the primary particles. In doing so we must take account of the deflection of the primary particles by the earth's magnetic field (Brunberg and Dattner 1953)). Finally, using the appropriate differential response curves (Webber and Quenby 1959), the variation of primary particle intensity can be converted into the corresponding variation for a particular component of the secondary radiation.

A calculation of this kind leads to an expected amplitude of $\sim 0.43\%$ for the nucleonic component at Huancayo with maximum intensity occurring at 13.30 hours local time. These values are in reasonable agreement with the experimentally observed amplitude of the nucleonic variation at Huancayo of 0.3% and the time of maximum at 12.15 hours (de Moura *et al.* 1959, private communication). This calculation of the daily variation to be expected from the solar field is in the nature of a preliminary estimate but it serves to show that there is no gross inconsistency between the observed daily variation and that expected on the basis of the present model.

§ 10. THE PROPAGATION TO THE EARTH OF PARTICLES ACCELERATED IN SOLAR FLARES

It is well known that occasionally particles with energies of a few gev can reach the earth from a solar flare and that in the early stages they do so in well defined impact zones. These impact zones soon disappear, however, leaving a slowly decaying flux of solar particles which is isotropic to a high degree. Cocconi *et al.* (1957) have shown that it is possible to account for this sequence of events if the photospheric magnetic fields can be extended by outward moving solar material into the form of long tongues which trap the solar particles and guide them to the earth.

In § 8 we have seen that distortion of the general field by outward-moving solar material can account for the Forbush decrease. In an extreme case

the general field may well be dragged out into a tongue with the same capabilities of guidance and storage as those envisaged in the model proposed by Cocconi *et al.* If we accept their picture we must conclude that the present model is in no way incompatible with the propagation characteristics for solar flare particles.

§ 11. SUMMARY

It appears from the calculations given above that the general characteristics of the cosmic-ray intensity modulation process can be explained in terms of a large-scale solar field of predominantly dipole character which contains small-scale irregularities acting as scattering centres and which is occasionally very seriously distorted by outward-moving solar material. In particular this model appears to be capable of giving a semi-quantitative account of the 11-year variation, the Forbush-type decrease and the solar daily variation. The 27-day intensity variation appears to be basically the same phenomenon as the Forbush decrease (Van Heerden and Thambyahpillai 1955) and consequently does not require separate treatment.

In fitting the field model to the experimental data on the 11-year cycle three adjustable parameters have been used. The first of these M_s , determines the strength of the field, assumed to be predominantly dipole, whilst the second, αK , is a measure of the probability of absorption by the sun of particles diffusing through the dipole field. The values of M_s lie in the range 1 to 3×10^{35} gauss cm³ and those of αK in the range 0.15 to 0.30. The third parameter, p_0 is related to the size and field strength of the scattering irregularities by the equation $p_0 = 300 B_i l$, and if we take $B_i \sim (M_s/r_e^3)$ and $p_0 \sim 1$ ev, as indicated by the cosmic-ray data, we find $l \sim 10^{11}$ cm near the earth. No particular care has been taken in the present paper to fit the experimental data with great accuracy but it would appear that if we are to attempt a more detailed description of the interplanetary field on the basis of cosmic-ray data we shall need greater observational accuracy on the primary spectrum variations than at present exists. If however, as suggested here, there exists a dipole like field with a strength of 3 to 8×10^{-5} gauss at the earth's orbit its detection by space probe magnetometer measurements cannot be long delayed.

ACKNOWLEDGMENTS

I am indebted to my colleagues at Imperial College and particularly to Professor P. Morrison, Dr. J. J. Quenby and Dr. W. R. Webber for many helpful and stimulating discussions on this subject.

REFERENCES

- ALFVEN, H., 1947, *Phys. Rev.*, **72**, 88; 1954, *Tellus*, **6**, 232.
 BABCOCK, H. D., 1959, *Astrophys. J.*, **130**, 364.
 BRUNBERG, E. A., and DATNER, A., 1953, *Tellus*, **5**, 135, 269; 1954, *Ibid.*, **6**, 73.

- COCCONI, G., GOLD, T., GREISEN, K., HAYAKAWA, S., and MORRISON, P., 1957, I.U.P.A.P. Cosmic Ray Conference, Varenna.
- EHMERT, A., 1959, I.U.P.A.P. Cosmic Ray Conference, Moscow.
- ELLIOT, H., HYND, R. J., QUENBY, J. J., and WENK, G., 1959, I.U.P.A.P. Cosmic Ray Conference, Moscow.
- FENTON, A. G., FENTON, K. B., and ROSE, D. C., 1958, *Canad. J. Phys.*, **36**, 824.
- JANOSSY, L., 1937, *Z. Phys.*, **104**, 430.
- KANE, E. O., SHANLEY, T. J. B., and WHEELER, J. A., 1949, *Rev. mod. Phys.*, **21**, 51.
- MORRISON, P., 1955, Proceedings of the I.U.P.A.P. Cosmic Ray Conference, Guanajuato.
- PARKER, E. N., 1958, *Phys. Rev.*, **110**, 1445.
- SINGER, S. F., 1957, I.U.P.A.P. Cosmic Ray Conference, Varenna, *Nuovo Cim.*, **5**, Suppl., 11.
- TREIMAN, S. B., 1954, *Phys. Rev.*, **93**, 544.
- WEBBER, W. R., and QUENBY, J. J., 1959, *Phil. Mag.*, **4**, 654.

Electron-microscopic Observations on Radiation Damage in Graphite†

By W. BOLLMANN

Battelle Memorial Institute, Geneva, Switzerland

[Received February 4, 1960]

ABSTRACT

Transmission electron microscopy has been applied to the study of lattice defects introduced by neutron irradiation of graphite. The dark-field technique has proved to be especially useful for this purpose. The technique of preparation is described and a tentative interpretation of the observations is given.

§ 1. INTRODUCTION

A GREAT deal of research work has been done on radiation damage in graphite, and this has been reviewed up to 1956 in various papers presented to the "First Conference on Peaceful Uses of Atomic Energy" (1956). An important problem is the mechanism of energy storage. The research work presented here, which is a part of a more extended research programme, has been undertaken with the aim of studying, by means of transmission electron microscopy, lattice defects introduced by neutron irradiation.

§ 2. TECHNIQUES

For the preparation of the microscope specimens, graphite scrapings, suspended in a water-alcohol mixture, were collected with a normal 200-mesh copper grid, which was covered by a collodion foil full of holes of about $1\ \mu$ diameter. Thus, thin graphite flakes could be placed over such holes without being backed by amorphous carbon from the collodion support. Another technique was similar to that used in taking a replica: a Triafol (acetobutyrate) foil wetted, and on its surface dissolved by acetone, was pressed on to a piece of graphite. After the foil had dried on the graphite, small particles stuck to it. A small piece of the Triafol was placed on the copper-collodion grid in a Soxhlet where the Triafol was dissolved in acetone. By this method very thin specimens could be obtained. A 'dry' method was used by dusting the graphite on to a sticky copper grid (the sticky grid was prepared by dipping a normal grid into a dilute benzene solution of 'Scotch-tape' cement). In this case the graphite flakes stuck out in all directions and very thin foils which are the most interesting ones, were found to be folded. The observations were made in a Siemens Elmiskop I at 80 kv.

† Communicated by the Author.

The pictures are taken in dark-field illumination in the 'light' of 1010 reflections selected by the aperture diaphragm, the illumination system being tilted to allow the selected beam to pass through the centre of the lenses, so that the full resolution of the microscope could be obtained. Dark-field, as compared with bright-field, has the advantage of much higher contrast because practically only those electrons which are diffracted contribute to the image, and the undeviated or diffusely scattered electrons are eliminated. Thus the influence on the image of the amorphous contamination caused by oil vapours is greatly reduced. In addition, only a small group from all the diffraction spots is selected, giving detailed information about a small component of the whole specimen. From a stack of layer-packets rotated with respect to each other, particular ones can be selected by this method. A comparison of fig. 4† and fig. 5 shows the difference in information obtained from the same area by dark-field and by bright-field illumination. In the case of the bright-field picture, none of the fine details of radiation damage can be seen. It is thought that for other applications also, dark-field transmission electron microscopy will markedly increase the amount of information obtainable, compared with the usual bright-field method. It is noteworthy, however, that even if only a part of a stack of graphite layers becomes visible by diffraction, the remainder of the material can have a harmful influence because of inelastic scattering, which introduces a chromatic error, with the effect of fading out fine details. The difference in concentration of fine details between the centre and the top of fig. 1 can be attributed to this effect. Normally, complicated moiré fringe systems are observed owing to relative rotation of superposed groups of graphite layers. The specimen has to be scanned in dark-field to choose areas with especially widely spaced moiré fringes, i.e. especially small rotation of superposed layers. The spacing of the fringe in fig. 1 corresponds to a rotation angle of $16.7'$.

§ 3. RESULTS AND DISCUSSION

The graphite studied was grade A reactor graphite irradiated by 10^{20} n/cm² at about 30°C. Figures 1 and 2 show the situation as it was observed in this material. Black and white dots with diameters ranging from 10 to 100 Å are visible. In another specimen of the same material small black and white loops can also be seen (fig. 3). Unirradiated graphite showed none of these dots: in this case only smooth moiré fringe systems were visible. Graphite which had been irradiated by 10^{17} n/cm² showed the same type of dots but at a much lower density.

To study the effects of annealing, specimens were examined of irradiated graphite (10^{20} n/cm²), heated in vacuum for 3 hours to different temperatures, i.e. 200°, 400°, 1000°, 2200°C. The strongest annealing effect is to be expected around 200°C as the differential energy release there is

† All figures are shown as plates.

a maximum (Woods *et al.* 1956). Figure 4 shows a specimen heated to 400°C. Both a certain coarsening of the loops and a decrease of the dimensions of the white and black dots are observed. In addition certain lines appear, which might be dislocations. Black and white dots were also observed in a specimen annealed at 1000°C but a specimen heated to 2200°C showed the same appearance as unirradiated material. The parallel annealing experiments of unirradiated graphite showed none of these features.

To obtain complete understanding of the nature of these dots, the contrast due to electron diffraction effects will have to be studied in detail. Tilting experiments have shown that the appearance of the dots can change from white to black and vice versa. The loops are reminiscent of those observed by Hirsch *et al.* (1958) in quenched aluminium, which are formed by vacancy clusters. Thus, as a tentative explanation, we suggest that the observed defects might be extrinsic and intrinsic stacking faults or groups of these corresponding to interstitial and vacancy clusters.

The difficulty in estimating the density of the dots lies in the fact that the thickness of the layer is not known. To produce the picture, fig. 1, at least three to five monatomic layers of graphite are needed, but probably there are more of them. If we suppose about 10 layers, 10^{20} n/cm², and a mean free path of the neutrons of about 3 cm, we observe about 5 to 10 times fewer dots than expected neutron impacts. Thus, the dots could originate from individual neutron impacts, if only the high-energy neutrons produce a visible effect. This does not mean that the dot should be at the place where the impact actually occurred, because it could mark the end of the path of an atom which has been hit, unless further rearrangement and annealing happened in the time between irradiation and observation.

Some of the properties of irradiated graphite, e.g. x-ray line broadening and differential energy release at 200°C, show a significant change for a dose of about 4 to 6×10^{20} n/cm² (Woods *et al.* 1956). At higher doses the x-ray lines broaden very strongly and the energy release on annealing is more and more displaced to higher temperatures at the expense of the release at 200°C. Looking at fig. 1 and considering a 6-times higher dose it becomes obvious that a significant part of the graphite lattice must be disturbed, or even destroyed, which would be compatible with the x-ray line broadening. So, the change from the low dose annealing, happening essentially at around 200°C to the high dose annealing which is displaced to higher temperature might correspond to the change from repairing an existing crystal lattice to nucleating a new one.

ACKNOWLEDGMENTS

The author wishes to thank Drs. P. B. Hirsch, J. Spreadborough, and G. Baroody for very helpful discussions, Miss R. Kollrack for the

preparation of the specimens and for partly developing the preparative technique, and AGIP Nucleare, Milan, for financing the research and for permission to publish this paper.

REFERENCES

- HIRSCH, P. B., SILCOX, J., SMALLMAN, R. E., and WESTMACOTT, K. H., 1958, *Phil. Mag.*, **3**, 897.
Peaceful Uses of Atomic Energy, Vol. 7, 1956 (New York : United Nations).
WOODS, W. K., BUYP, L. P., and FLETCHER, J. F., 1956, *Peaceful Uses of Atomic Energy*, Vol. 7 (New York : United Nations), p. 455.

Cross-tie Walls in Thin Permalloy Films†

By M. PRUTTON

International Computers and Tabulators Ltd., Research and Design Division,
Gunnels Wood Road, Stevenage, Herts

[Received February 15, 1960]

ABSTRACT

A solution of the non-linear differential equations of a domain boundary is reported for the case of a thin film in which the demagnetizing energy associated with the magnetization at right angles to the plane of the film is included. The solution leads to a proposal for the internal structure of the cross-tie wall which has been observed in thin films of permalloy. The effects of the magnetic poles in this model of a cross-tie wall are discussed and compared with observations. The energy density of the wall is suggested to be about 10 ergs/cm² and the correlation between cross-tie spacing and length is derived.

§ 1. INTRODUCTION

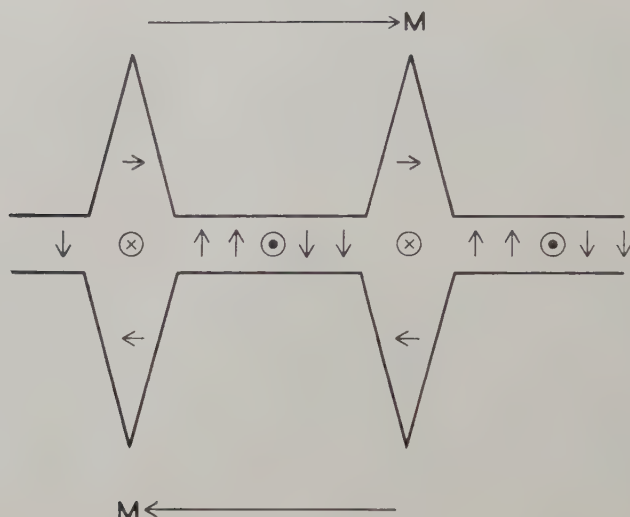
IN order to understand coercivity effects in thin films when the reversal process is one of domain wall motion it is important to have detailed knowledge of the energies of the domain boundaries. Néel (1955) has described how the energy associated with a Bloch wall rises as the thickness of a film decreases because of the increasing magnetostatic contribution to the wall energy arising from the poles on the surface of the film. When this magnetostatic energy becomes sufficiently great it is energetically favourable for the magnetization M in the boundary region to rotate in the plane of the film. Such a wall is known as a Néel wall and in permalloy films of zero magnetocrystalline anisotropy and induced uniaxial anisotropy of 1500 ergs/cm³ it occurs below a thickness of about 300 Å and decreases in energy with decreasing film thickness. Further, the Néel wall is much thicker than a Bloch wall for the same film thickness because of the low value of the induced anisotropy which provides the force tending to restrict the wall thickness.

Huber *et al.* (1958) have observed a 'cross-tie' structure domain wall in thin films of permalloy and have proposed a model for the magnetization distribution within the wall. Figure 1 is reproduced from their paper and it shows how their model includes both Néel and Bloch like segments.

† Communicated by the Author.

The object of this paper is to describe some calculations made to try to derive a cross-tie wall from a perfectly general set of equations for a domain wall.

Fig. 1

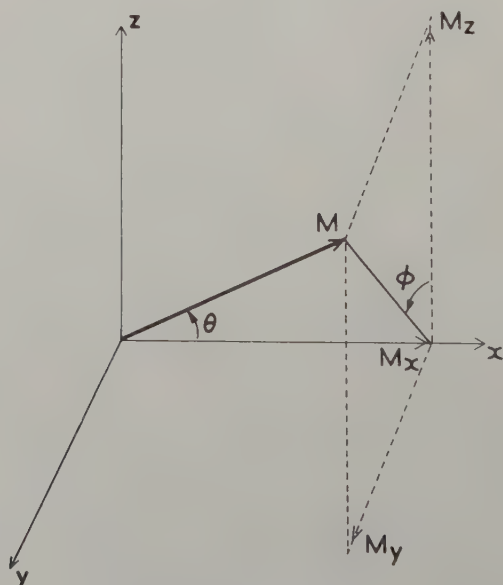


Huber, Smith and Goodenough's model for the cross-tie wall.

§ 2. SOLUTION OF THE WALL EQUATIONS

By describing the behaviours of the amplitude-constant magnetization vector \mathbf{M} with two angles (θ, ϕ) defined in fig. 2 and in the presence of a

Fig. 2



The definition of coordinate angles.

general anisotropy, $K(\theta, \phi)$ which can be chosen to include magnetostatic, magnetoelastic and magnetocrystalline terms Harrington (1958) has shown that the conditions to be satisfied by a static boundary in the absence of an applied field and in a material with exchange constant A are

$$2A \nabla^2 \theta = 2A \sin \theta \cos \theta (\nabla \phi)^2 + \partial K / \partial \theta, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$2A \sin^2 \theta \nabla^2 \phi = -4A \sin \theta \cos \theta (\nabla \theta \cdot \nabla \phi) + \partial K / \partial \phi. \quad . \quad . \quad . \quad (2)$$

Harrington derived these equations in a semi-classical manner by formulating a Langrangian of the system and then using a variational treatment to obtain a set of Euler-Lagrange equations. It was from these Euler-Lagrange equations that eqns. (1) and (2) were deduced.

In a thin permalloy film the total anisotropy K can be represented by two terms:

(i) A term $K \sin^2 \theta$ expressing the fact that such films show an induced uniaxial anisotropy and

(ii) A term proportional to the square of the component of magnetization at right angles to the plane of the film which occurs because of the magnetostatic energy associated with magnetic poles at the surface of the film. Thus the total anisotropy term can be written as

$$K(\theta, \phi) = K \sin^2 \theta (1 + b \cos^2 \phi) \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where b is a constant determined by the magnetostatic energy density of a thin film magnetized uniformly perpendicular to its plane.

The solution of eqns. (1) and (2) for a given set of boundary conditions is not straightforward and in order to proceed from the $K(\theta, \phi)$ in (3) above let us choose to find a solution in which θ varies normal to the wall and ϕ varies along the wall so that the derivatives of the two angles are orthogonal and $\nabla \theta \cdot \nabla \phi = 0$.

Thus the two equations become:

$$A \partial^2 \theta / \partial y^2 = A \sin \theta \cos \theta (\partial \phi / \partial x)^2 + K \sin \theta \cos \theta (1 + b \cos^2 \phi), \quad (4)$$

$$A \partial^2 \phi / \partial x^2 = -bK \cos \phi \sin \phi. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Equation (5) may be directly solved giving

$$x - x_0 = \sqrt{\left(\frac{A}{bK}\right)} \log \tan (\phi/2 + \pi/4) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

inserting $(\partial \phi / \partial x)^2$ in (4) and solving yields the result:

$$y - y_0 = \sqrt{\left(\frac{A}{K(1 + 2b \cos^2 \phi)}\right)} \log \tan (\theta/2). \quad . \quad . \quad . \quad . \quad (7)$$

Equations (6) and (7) together show that the wall thickness (y direction) varies with the value of ϕ and hence with the value of x . A plot of y at $\theta = 175^\circ$ against x is shown in fig. 3 using the following values

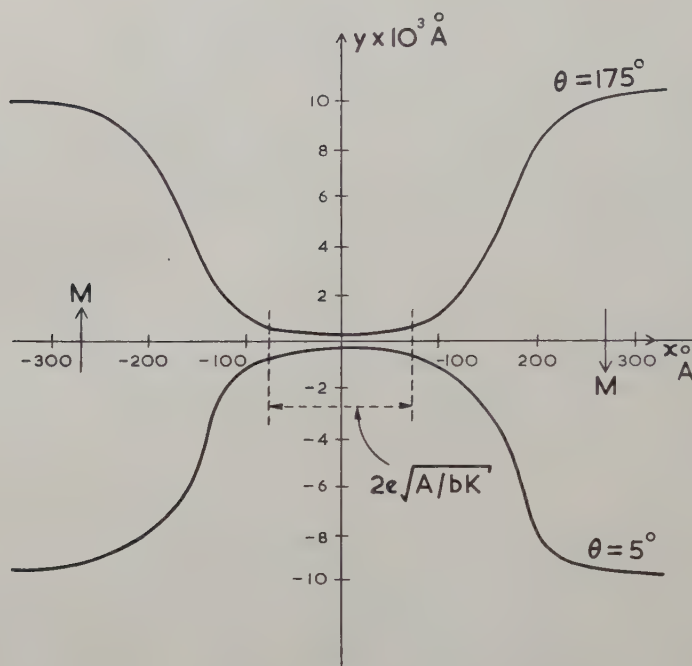
$$A = 10^{-6} \text{ ergs/cm}$$

$$K = 1500 \text{ ergs/cm}^3 \text{ (i.e. } H_k \sim 4 \text{ oersteds),}$$

$$b = (4\pi M^2 / K) = 8.5 \times 10^3.$$

This value of b has been used because the energy associated with the demagnetizing field normal to the plane of the film is $4\pi M^2$.

Fig. 3



The variation of wall thickness for $\theta = 5^\circ$ and $\theta = 175^\circ$.

It is clear that the solution represented by eqns. (6) and (7) is not cyclic in x so that it cannot be the solution for a cross-tie wall. However, the following points are to be noticed.

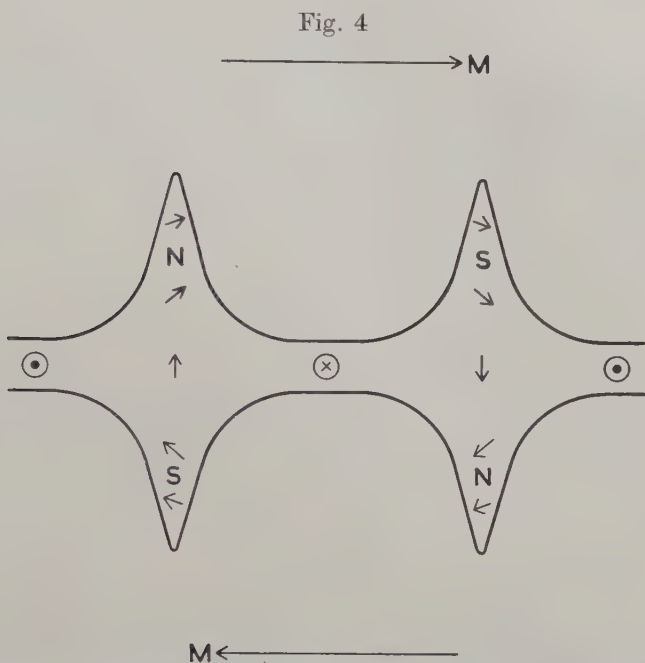
(i) This solution contains a Bloch segment and two Néel segments and so is a compromise between pure Bloch and Néel walls in the same way as the original interpretation by Huber, Smith and Goodenough of the cross-tie wall.

(ii) Using the constants above, the distance between the centre of the Bloch segment and the Néel segment is about 300 Å which is a distance very short compared with the diameter of a typical film. The magnetostatic energy associated with the free poles in the Néel segments has not been taken into account in the calculation above and it is likely that the Néel segments will be reduced in length in order to minimize this energy. Further, there will be a magnetostatic interaction between the poles in adjacent Néel segments. This will be discussed later.

(iii) The wall thickness in the Néel region is of the order of 15 000 Å which compares reasonably with the cross-tie width as shown in the paper by Huber *et al.* and in a paper by Moon (1959).

(iv) With increasing film thickness the value of b falls and both the length and the width of the Bloch segment increase so tending to a simple Bloch wall. The wall thickness in the Néel segment decreases with increasing uniaxial anisotropy K as is expected.

It is suggested that the solution above could in fact represent one cycle of a cross-tie wall and the reason that a cyclic solution has not been found is associated with the fact that the original formulation of the domain wall equations did not take into account the energetic consequences of a non-constant value of $\text{div } \mathbf{M}$ in the wall. These magnetostatic interactions will be dependent upon the film thickness and will be important in determining the width and spacing of the Néel segments. In fig. 4 is drawn a model for



Proposed model for the magnetization distribution within a cross-tie wall.

the proposed structure of a cross-tie wall based upon the solution derived above. The principal difference between this model and that of Huber *et al.* is that the wide sections of the wall are Néel segments and do not contain a magnetization vector at right angles to the plane of the film at the centre of the wall. This model is still consistent with the shift observed in a powder pattern when a field is applied in a plane normal to the surface of the film.

The magnetic colloid will reveal a cross-tie type of wall structure for this magnetization distribution because it will lie in the places where the magnetic field at the surface of the film is high. Thus it will collect along the length of the wall because here the magnetization intersects the film

surface and there are magnetic poles, and along the cross-ties because here there is a stray field linking the poles in the two halves of Néel segment on each side of the wall.

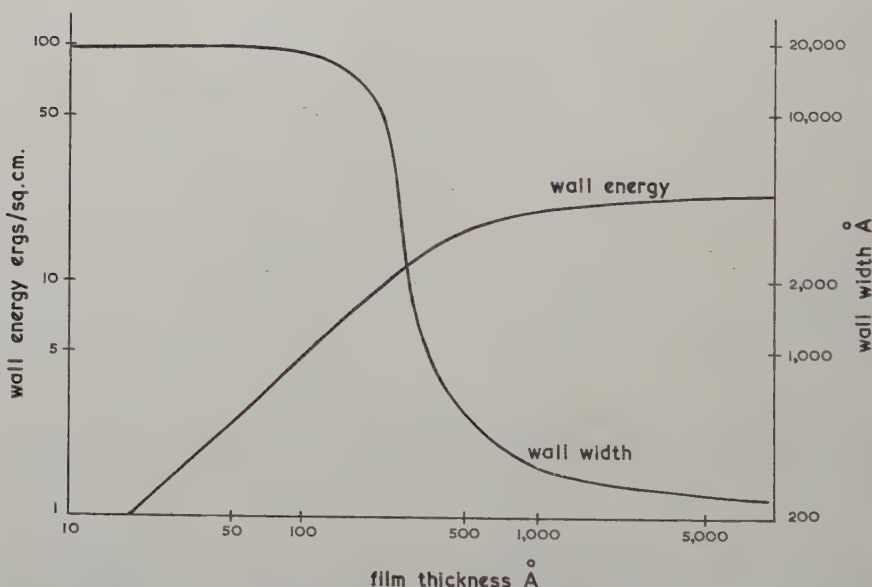
§ 3. THE EFFECTS OF MAGNETOSTATIC INTERACTIONS

Before attempting to evaluate the effects of magnetostatic interactions in the cross-tie wall it is convenient to consider the thickness and energy of the simple Néel type of boundary. If the energy density and width of a wall in the bulk material are written as σ_0 and a_0 respectively then the energy density in a Néel wall of width a in a film of thickness D can be written (Néel 1955) as:

$$E = \frac{\sigma_0}{2} \left(\frac{a}{a_0} + \frac{a_0}{a} \right) + \frac{2\pi a D M^2}{a + D}.$$

The equilibrium wall energy and width for a film of given thickness can be found from this equation by putting $\partial E / \partial a = 0$. The variation of E and a with film thickness is shown in fig. 5. The variation in the width of the

Fig. 5



The thickness and energy density of Néel wall in permalloy film using $\sigma_0 = 0.28$ ergs/cm², $a_0 = 2 \times 10^{-4}$ cm, $A = 10^{-6}$ ergs cm and $K = 10^3$ ergs/cm³.

wall with changing film thickness is attributed to the fact that as the film thickness increases the area of free poles in the wall increases and as these are of opposite sign in each half of the wall they attract each other increasingly and the wall becomes thinner.

Next it is instructive to consider the wall shown in fig. 4 when the free poles in the Néel segments are regarded as being at a point distance p from the axis of the wall and the cross ties are separated by a distance q . It is

easily shown by considering nearest neighbours and magnetostatic effects alone that the wall is stable when

$$256p^6 = (q^2 + 4p^2)^3$$

i.e.
$$q \sim 1.5p.$$

The dimension p is determined primarily by the free poles in the Néel segments alone, consequently as the film thickness decreases the dimensions p and q both increase. This is consistent with the observations of Rubinstein and Spain (1959) who report that the spacing between cross ties increases as the films thickness decreases and that the larger the spacing between cross ties the longer is the cross tie.

§ 4. THE WALL ENERGY

The sum of the exchange energy and the anisotropy energy at a point (θ, ϕ) in the wall is

$$f = 2K \sin^2 \theta (1 + 2b \cos^2 \phi).$$

Neglecting the magnetostatic contributions the wall energy per unit length can then be written

$$\sigma_l = \int_{-\infty}^{\infty} f dy = \int_0^{\pi} f(dy/d\theta) d\theta.$$

Thus

$$\sigma_l = 4\sqrt{[AK(1 + 2b \cos^2 \phi)]}.$$

An approximate term for the average wall energy density σ_w can then be derived by assuming that the solution above is one for a cyclic wall and averaging the energy σ_l over one cycle.

Thus

$$\sigma_w = \frac{\int_{-\pi/2}^{\pi/2} \sigma_l(dx/d\phi) d\phi}{C\sqrt{(A/bK)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where C is a constant which allows for the fact that the cycle length is in fact longer than $\sqrt{(A/bK)}$. It can be seen from fig. 3 (where the value of $\sqrt{(e^2 A/bK)}$ is marked) that a reasonable value for C is about 2.5. Equation (8) can then be expressed as

$$\sigma_w = \frac{8M}{C} \sqrt{(2\pi A)} \int_0^1 \sqrt{\left(\frac{a^2 + y^2}{1 - y^2}\right)} dy$$

where $a = 1/2b$. The integral is a complete elliptic integral of the second kind which is equal to unity because $b \gg 1$, therefore

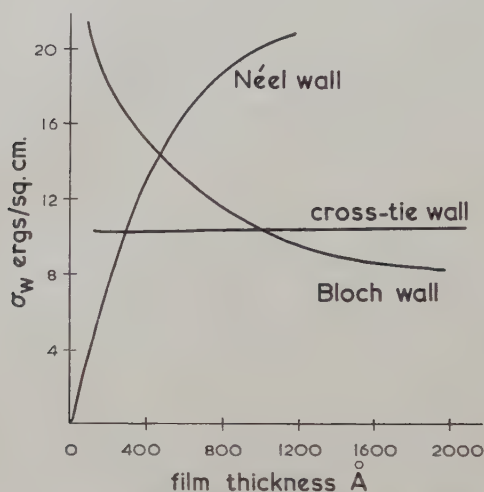
$$\sigma_w \sim 3M \sqrt{(2\pi A)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

For the values of M and A used above this means that $\sigma_w \sim 10$ ergs/cm² and is independent of film thickness and uniaxial anisotropy K when the magnetostatic contributions are neglected. This latter result is reasonable

because in this case the energy balance is between the exchange A and the demagnetizing field $4\pi M$.

Using Néel's data for Bloch and Néel walls fig. 6 has been plotted showing the variation of wall energy with film thickness for the three types of wall.

Fig. 6



Plots of wall energy density *versus* film thickness for Bloch, Néel and cross-tie walls.

It is clear that a cross-tie wall of energy density 10 ergs/cm² can only be expected in thin permalloy films in the thickness range 200–1000 Å. This is in substantial agreement with the work of Methfessel *et al* (1959) who report that they can observe cross-tie walls in thin permalloy films only in the thickness range 400–900 Å.

§ 5. CONCLUSIONS

Although it has not been possible to derive a cross-tie wall directly from the general wall equations a solution describing a wall half-way between a Bloch and a Néel boundary has been found. This solution appears to be similar to one-cycle of a cross-tie wall and makes possible the presentation of a more plausible model for such a wall than that suggested by Huber, Smith and Goodenough. The energy balance in this wall is between the exchange energy and the demagnetizing energy contained in the component of the magnetization at right angles to the plane of the film. The spacing and lengths of the cross-ties are determined by the magnetostatic interactions between the cross-ties arising because they are Néel type regions and so contain magnetic poles. An approximate calculation of the energy of this wall reveals that permalloy films should only exhibit cross-tie walls when between 200 and 1000 Å thick. This is in fair agreement with observation.

ACKNOWLEDGMENTS

The author would like to thank H. Rubinstein of Laboratory for Electronics Inc., Boston, Massachusetts for a stimulating correspondence on this problem. He would also like to thank the Directors of International Computers and Tabulators Ltd., for permission to publish this paper.

REFERENCES

- HARRINGTON, J. V., 1958, *M.I.T. tech. Rep.*, Lincoln Laboratory No. 166, January.
- HUBER, E. E., SMITH, D. O., and GOODENOUGH, J. B., 1958, *J. appl. Phys.*, **29**, 294.
- METHFESSEL, S., MIDDLEHOEK, S., and THOMAS, H., 1959, Paper 122, Detroit Conference on Magnetism and Magnetic Materials; 1960, *J. appl. Phys.*, **31**, 302 S.
- MOON, R. M., 1959, *J. appl. Phys.*, **30**, 825.
- NÉEL, L., 1955, *C. R. Acad. Sci., Paris*, **241**, 533.
- RUBINSTEIN, H., and SPAIN, R. J., 1959, Paper 123, Detroit Conference on Magnetism and Magnetic Materials; 1960, *J. appl. Phys.*, **31**, 306 S.

The Generation of Vacancies in Metals†

By R. S. BARNES

Atomic Energy Research Establishment, Harwell, Berkshire

[Received January 7, 1960]

ABSTRACT

The regions where injected helium atoms first appear as gas bubbles are those where thermal vacancies originate. Microscopic examination of metals, injected with helium by using them as targets for energetic alpha-particles, reveals grain boundaries as the principal suppliers of vacancies. Detailed examination distinguishes those boundaries which generate vacancies from those which merely conduct them. The behaviour and nature of these boundaries and other vacancy sources are discussed.

§ 1. INTRODUCTION

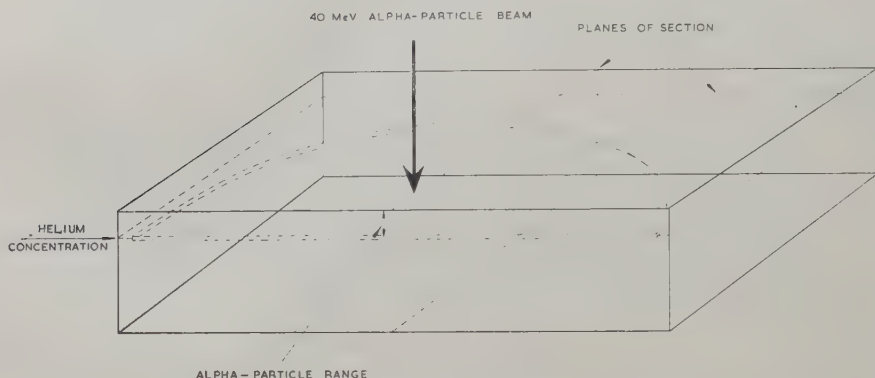
SINCE the concept of the vacant lattice site (or vacancy) was introduced by Frenkel, it has become increasingly clear that this point defect plays an important role in many physical processes of crystalline solids; particularly at high temperatures where the equilibrium vacancy concentration rises steeply, but even at low temperature when non-equilibrium vacancy concentrations have been introduced by quenching, deformation, or by atomic particle bombardment. A recently developed technique (Barnes *et al.* 1958) enables the sources of vacancies to be observed. These observations and some of the deductions from them will be given in this paper.

Mono-energetic alpha-particles from a cyclotron have been used to bombard a metal surface and inject a layer of helium atoms beneath the surface at a depth equal to the range of the alpha-particles in the metal (see fig. 1). The helium atoms in this layer, which is normally quite thin ($\sim 10^{-3}$ cm), are unable to migrate once their initial momentum is spent unless the temperature of the metal is sufficiently high. If the temperature is below this value the helium atoms remain dispersed in 'solution' in the metal, the concentration steadily increasing far beyond the solubility limit as the alpha-particles continue to be fired in. High concentrations of helium can be retained in solution in this way. On heating such a bombarded metal the helium atoms become mobile at some temperature and precipitate from solution with the ultimate formation of gas bubbles. Gas bubbles have been observed on heating copper (Barnes *et al.* 1958), aluminium (Barnes 1959), beryllium (Barnes and Redding 1959) and uranium which has been injected

† Communicated by the Author.

with helium. Some study has also been made of gas inserted by other means into the metals (Ells and Perryman 1959, Mykura 1959, Rich *et al.* 1959). Figure 2† is a photomicrograph of a sectioned copper single-crystal injected with helium by bombardment with 38 mev alpha-particles and then heated to 700°C for 1 hour. The bombarded surface is seen at the top of the figure,

Fig. 1



The injection of helium atoms by bombardment with alpha-particles.

and the layer containing the helium bubbles can be clearly seen at the bottom‡. It is observed that in the first instant only those helium atoms accessible to vacancies precipitate; only the periphery of the layer containing bubbles. This is the region immediately accessible to vacancies from outside the layer. After shorter times (or lower temperatures) the bubbles are even more marginal, and as the heating progresses the vacancies penetrate the layer further until the whole of the layer is filled with bubbles. These observations show that there are no vacancy sources *within* the helium layer in single crystals of copper. By heating for a time insufficient to allow the complete precipitation of the helium it is possible to outline with bubbles, the sources of the vacancies necessary for the process, and so observe these sources.

§ 2. VACANCY SOURCES

2.1. *Free Surface*

That the free surface of the metal is a source of vacancies is demonstrated by fig. 3. This shows that although the heat treatment has been insufficient to fill the helium layer with bubbles, near the free surface (lying on the left of the figure), the layer is filled. (The free surface has been covered with an

† Figures 2–11 are shown as plates.

‡ In this sample the individual bubbles cannot be resolved, yet the layer can be seen, mainly due to its lower reflectivity than the rest of the electro-polished surface.

electro-deposit of copper after the heat treatment to preserve flatness of the micrographed surface during polishing.) Vacancies have apparently formed at the free surface during heating, and have migrated inwards a distance demarked by the limit of the bubbles. The shape of this demarcation line must be an index of the concentration of the helium in the layer. If a given number of vacancies per helium atom is necessary before bubbles can form, then the penetration of the vacancies is inversely proportional to the helium concentration. If this is so the shape suggests that the helium is assymmetrically distributed about the mean range of the alpha-particles, and this is borne out by the distribution of the bubbles once they have formed.

Evidence that the free surface is a source of vacancies is provided also at each heat treatment, for it is only after a sufficiently thick layer has been removed from the surface that an unbubbled region is seen. For instance, fig. 4 shows the same area (*a*) after a thin surface layer has been removed and (*b*) after much more has been removed, when the vacancies from the original surface are not contributing to the helium bubbles.

2.2. Grain Boundaries

Vacancies are also generated at grain boundaries within the helium layer, Figure 4 shows this in the same piece of copper as that seen in fig. 3, but after heating at 825°C for 1 hour, and the penetration of vacancies into the layer is very similar, suggesting that the grain boundary supplies vacancies at the same rate as the free surface. Both figs. 3 and 4 were of copper which had been well annealed (heated to 1050°C for 48 hours *in vacuo*) before the helium was injected. Other samples of copper with a lower temperature of anneal had much smaller grain size. This material showed similar behaviour at the grain boundaries (see fig. 5), but due to the more numerous sources the vacancies were more plentiful. Photographs of samples sectioned in a plane through the centre of the band and parallel to the bombarded surface (parallel sections) are shown in figs. 5 and 6. These show clearly that most of the grain boundaries are supplying vacancies to the helium in the surrounding metal.

2.3. Twin Boundaries

There are many coherent annealing twin boundaries shown in figs. 4, 5 and 6. These are never surrounded by bubbles and thus are never vacancy sources. As such a coherent twin boundary is not surrounded by bubbles, even where it intersects a grain boundary which is, suggests that it is also unable to *conduct* vacancies. However, a *non-coherent* twin boundary, although normally free of bubbles, frequently is surrounded by them if it connects with a grain boundary, suggesting that such a boundary is conducting vacancies from the true vacancy source. Not all non-coherent boundaries do this, e.g. fig. 6 shows two non-coherent boundaries (NC) of similar nature, both intersecting bubbled regions but only one of these is bubbled itself.

There are certain twin boundaries which invariably contribute vacancies to their surroundings. These are boundaries separating two individual annealing twins. In every such case the boundary contributes vacancies equally as well as a normal grain boundary. Two examples of boundaries (one being very short) between the two twins T_1 and T_2 of the parent grain P are shown in figure 6.

2.4. *Point Sources*

In addition to the vacancy sources already described there are point sources which give rise to small circular regions of bubbles (see, for example, fig. 5). These point sources can in a few cases be identified as being due to voids within the metal. Figure 7 shows two photomicrographs of a void in aluminium. The first (*a*) is of a section almost through the centre of the void and (*b*) is of a subsequent section cutting the void below its centre. Such a void would be able to contribute vacancies to its surroundings until it was completely filled in by the aluminium atoms deposited there. The bubble free zone immediately surrounding the void is presumably due to these deposited atoms.

Inclusions which at the heating temperature are surrounded by a gap due to differential expansion should behave as a void, but due to the smallness of the free volume available the number of vacancies generated there should be limited. Figure 8 illustrates this behaviour at an inclusion (I) in copper. The inclusion is surrounded by a very narrow circle of extremely small bubbles ($\sim 300 \text{ \AA}$ in diameter) which can only just be discerned in the figure but have been revealed by electron micrographs of replicas from this region. In comparison the void[†], which lies very close to the inclusion in fig. 8, has supplied sufficient vacancies for bubbles to have formed a large distance from it.

Other point sources, which do not contain a helium free zone (see the electron micrograph of a replica in fig. 9) are less well understood. As can be seen, there are very few in material which has been well annealed (figs. 2, 3 and 4) but they are very numerous in material heated sufficient only for recrystallisation before injecting the helium (fig. 5). A piece of copper (that shown in fig. 6) which had been well annealed was cold-wrought and then recrystallized before injecting it with helium. The material was then found to contain point sources, but these were fewer in number and more irregularly sited than those seen in the rolled material treated identically, but not annealed at a high temperature (cf. fig. 6 with 5). It is unlikely that these point sources are due to voids, as it is difficult to believe that voids could be reintroduced merely by working the material.

2.5. *Dislocation Lines*

On no occasion in any of the material studied have dislocation lines themselves been identified as supplying vacancies to the helium atoms. The

[†] The void has permitted a greater penetration of those alpha-particles which have passed through it, as is shown by bubbles lying beneath the void which have prevented locally the advance of the grain boundary running along the base of the gas layer.

well-annealed material contains $\sim 10^6$ dislocation lines cm^{-2} and the recrystallized material many more, but in neither case are bubbled regions observed on a scale as fine as this. Samples have been purposely bent or cold-worked in other ways, but although slip occurred in the region containing the helium atoms, the subsequent precipitation of the helium followed the normal pattern for the undeformed material, despite the inevitable introduction of dislocations into the helium layer.

§ 3. DEPOSITION OF ATOMS

One consequence of the generation of vacancies is that there must be a deposition of atoms at the source. This deposition arises from the manner in which vacancies migrate and should be observable. A vacancy migrates by exchanging places with a neighbouring atom, so that although each atom only moves one jump, there is in effect always a counter flow of atoms in equal number to the flow of vacancies. Thus the generation of a vacancy at a source and its subsequent migration to a helium bubble is equivalent to the migration of an atom from the bubble to the source, where it is deposited. Thus for each vacancy generated at a source an atom of the matrix is deposited there.

The helium bubbles, which are inert markers in the crystal, will move away relative to the source as these atoms are deposited. Thus at each source there should be a region free of bubbles containing as many atoms of the matrix as there are vacancies generated. This phenomena can be observed for instance in fig. 7 (*a*) and (*b*) which shows a zone free of bubbles, at the surface of the void†. The width of the zone free of bubbles is relatively much wider in the lower section (fig. 7 *b*) as would be expected. Eventually the void can be completely filled and will then cease to be a source of vacancies as occurs at the inclusion shown in fig. 8. The majority of point sources do not contain such a region free of bubbles suggesting that they operate in a different way.

Figure 10 (*a*) is a shadowed two-stage carbon replica of the electro-polished surface of a region in the helium filled layer in copper after heating for two hours at 800°C . The individual bubbles are easily resolved and each casts a shadow. The diameter of the bubbles can be estimated from the width of these shadows, but these may over-estimate due to enlargement by the etchant. Figure 10 (*b*) shows a thin film of copper ($\sim 3000 \text{ \AA}$ thick) viewed in transmission in the electron microscope. The bubble size is here truly represented as the majority of them are completely embedded in the copper (diameter $3 \times 10^{-5} \text{ cm}$). Both these photographs show grain boundaries running across them and at each there is a region free from bubbles, similar to that observed at the void in fig. 7.

If the explanation for this free zone is correct then the volume of the bubbles formed per unit area of grain boundary should be equal to the volume

† No such clear zone can be seen at the free surface shown in fig. 3. This may have been etched away during the preparation for the electro-deposition, or smeared out by surface diffusion.

of the material deposited upon that area of the boundary. Rough estimates from fig. 10, where the average bubble radius is taken to be 1.5×10^{-5} cm and the number of bubbles $\sim 10^{13} \text{ cm}^{-3}$ show that the volume of the deposit is about 10^{-4} cm^3 and the equivalent bubble volume $1.3 \times 10^{-4} \text{ cm}^3$; agreement well within the accuracy. The number of vacancies generated at each atom site of the grain boundary is thus 3×10^3 during the two hours heating at 800°C and the deposited atoms produce a local volume increase around the boundary which in the above example is about 11%.

Not all grain boundaries show this effect, suggesting that only some of the grain boundaries (those with helium-free zones) are really acting as sources the rest are merely conducting vacancies into the crystal from elsewhere. Some local diffusion of the helium to those boundaries surrounded by bubbles occurs as some small bubbles can always be seen on the boundaries themselves. The helium free region cannot be caused by diffusion of the helium however, as this would occur at all boundaries equally.

In general the grain boundaries which are the most copious vacancy sources are those which lie in, or nearly in, the plane of the helium layer. This behaviour can be understood when it is realised that if a grain boundary is to be a true source of vacancies it must have atoms deposited upon it, and the region must *expand* laterally. A grain boundary lying perpendicular to the helium layer will be heavily restrained by the metal above and below the layer and such expansion will be difficult. Grain boundaries lying in the plane of the layer will have no such restraint.

Boundaries which do not have this helium free region frequently contribute fewer vacancies into the surrounding crystal as the distance from the real source of vacancies increases (e.g. some boundaries shown in fig. 5). This adds to the evidence that they are only conducting vacancies from this source.

§ 4. DISCUSSION

The observations show that the first bubbles to form surround both the true vacancy sources and also those features which are merely conducting vacancies from these sources. A source can be distinguished from a vacancy conductor by the region of precipitated metal atoms, devoid of helium bubbles, adjacent to it. The vacancy sources and conductors will be discussed separately.

4.1. Vacancy Sources

The free surface is the most straightforward source of vacancies. On a clean surface there will be no restraint to reduce the normal atomic fluctuations which occur. A surface film, e.g. an oxide layer, may retard such a vacancy source, but if it is not a coherent one the effect should be slight.

The situation at a grain boundary is similar to that at a free surface with the added complication caused by the superimposed crystal of a different orientation. The inherent structure of a boundary is expected to determine its ability as a vacancy source. If there is a space in the boundary of

approximately an atomic size then there would be little hinderence to the generation of a vacancy. Boundaries having little free space, i.e. nearly coherent boundaries, would not be good sources. However, not only must there be spaces in the boundary large enough for extra atoms to be forced in but spaces must then form elsewhere in the boundary for the process to continue.

A hydrostatic pressure (P) upon a sample generating vacancies would necessitate an extra amount of work Pa^3 (where a^3 is the volume of an atom) to be performed in generating a vacancy. A pressure of more than 10^4 atmospheres must be applied before this extra work ($\sim 7 \times 10^{-2}$ ev) is comparable with that needed to create a vacancy at a free surface (~ 1 ev). If such a pressure was applied then the number of vacancies emanating from each and every source would be limited. Such high pressures cannot be maintained within these samples, so that local pressure, e.g. those caused by unequal vacancy production at different boundaries, cannot influence the emanation of vacancies from a boundary.

However, internal pressures, although unable to prevent the formation of bubbles, can change the behaviour of a boundary from that of a source to that of a conductor by forcing the atoms which are normally deposited on the boundary to flow along it to some other boundary (not subjected to this pressure) where they are deposited. A lateral pressure of about 40 atmospheres (600 p.s.i.) is estimated to be sufficient to cause this flow of atoms out of a boundary in a typical experiment. Such pressures can be attained and maintained, and explain why boundaries lying in the plane of the helium layer contain the zone free of bubbles near to them and those lying perpendicularly do not. Such pressures around a cylindrical or spherical source would also be expected to force the atoms to escape by any possible path. If no such path existed then the deposited atoms would build up a pressure sufficient to cause the surrounding material to yield.

Thus external pressure, or pressures imposed by the material itself, can induce a vacancy source to act as a conductor. But this cannot explain why a coherent twin boundary is never a source of vacancies. This must be due to the nature of the boundary itself.

4.2. Vacancy Conductors

Certain crystal defects are unable to generate vacancies but can conduct them†. Such conductors are distinguished from true sources by the absence of an intervening region of deposited metal atoms between them and the helium bubbles. There is no change in volume in such a region as the vacancies are conducted in from their true source where the actual volume increase occurs. The conduction process is thus unaffected by restraining pressures except in so far as the source itself is affected.

† Strictly, the concept of a vacancy in a boundary is meaningless because of the absence of a simple crystallographic arrangement there. But the end effect of the diffusion of atoms (not necessarily a particular atom) along the boundary is equivalent to the diffusion of a vacancy along the boundary.

A grain boundary which appears to be a good source is sometimes converted into a poor conductor where one of the adjacent grains is twinned (fig. 11) and occasionally a change in boundary direction prevents a boundary from even conducting vacancies. These effects cannot be caused by local pressure changes and clearly show the importance of the nature of a boundary upon its behaviour.

As a conductor can only conduct a limited number of vacancies from the true source, and as the vacancies are 'pumped' away into the helium bubbles in the surrounding crystal, the supply becomes exhausted at a distance from the real source which is dependent upon the relative diffusion rates of vacancies along the conductor and through the surrounding crystal. In small grain material the penetration of vacancies from the source boundaries into the surrounding crystal is the same for all, but in large grain material the distance is very variable. This is because in the latter case the conduction paths are much longer so that exhaustion of the supply occurs more frequently.

Occasionally a conducting boundary is seen in a parallel section, where the bubbled region tapers to a point, the vacancy supply being exhausted there (see fig. 5). In some of these cases the penetration of vacancies from the adjacent vacancy source is more than the width of the unbubbled region of the helium layer, i.e. more than the penetration in a direction perpendicular to the photograph. Thus in these particular boundaries the conduction of vacancies is anisotropic. Vacancies conducted along the boundary from outside the layer are unable to travel as far as those conducted in the plane of layer. In other cases the exhaustion occurs in less than the width of the unbubbled region and here the boundary is probably a poor conductor in all directions.

The circular regions of bubbles observed in figs. 5 and 6 suggest that vacancies are emanating from a point. They cannot arise from voids which were originally present but are now completely filled as there is no central region devoid of helium bubbles (see fig. 9). The complete absence of such central regions also demonstrated that these spheres of bubbles are not caused by point *sources* of vacancies. Vacancies must be supplied to the region and then radiated from the central point. As these regions do not in general lie on any observable boundary the vacancies must be conducted in via some unseen defect, presumably dislocation lines. Certainly the number of circular regions is very much smaller and their size more non-uniform in well annealed material (in fact none can be seen in figs. 2, 3 and 4 where the number of dislocation lines is $\sim 10^6 \text{ cm}^{-2}$). In this material the number of circular regions is only $\sim 10^4 \text{ cm}^{-2}$ ($r \sim 2 \times 10^{-4} \text{ cm}$), i.e. the number of spherical regions is $\sim 2 \times 10^7 \text{ cm}^{-3}$, whereas the number of dislocation nodes must be $\sim 10^9 \text{ cm}^{-3}$, almost two orders of magnitude greater. Thus if the dislocation nodes are the originating points, only a few of them are operating. The regions might originate upon some special dislocation configuration which can climb and thus supply vacancies itself.

4.3. Dislocation Lines

It is surprising that the majority of the dislocation lines which are known to exist in the metal are unable to climb and thus supply vacancies to the helium atoms and permit precipitation as bubbles. Dislocations of both edge and screw character are known to be sinks for supersaturated point defects and it seems reasonable that they should behave reversibly when the vacancy concentration is depleted. The explanation may be that the dislocations are effectively locked by the helium atoms lying upon them and are thus unable to *climb*. Certainly the material in the layer containing helium is much harder (the microhardness is approximately doubled) presumably because the helium atoms impede *slip* of dislocation lines.

§ 5. STRUCTURE OF BOUNDARIES

A coherent twin boundary, which is never either a source or a conductor of vacancies, is a particularly simple boundary, lying on a (111) plane. The diagram (fig. 12) shows the atom positions of a (111) plane as dots. The

Fig. 12

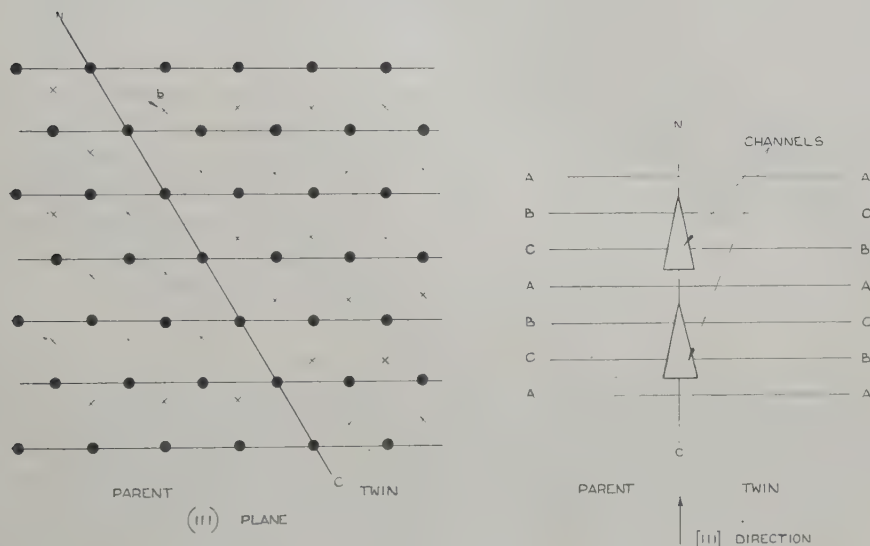


Diagram of the atomic arrangement at a non-coherent boundary NC in a face-centred-cubic crystal.

plane of atoms above this is represented as crosses. The right hand side can be seen to have suffered a twinning shear, i.e. the passage of a partial dislocation with Burgers vector b on each plane. If the successive (111) planes in a perfect crystal are labelled ABCABC then the fault at the twin plane can be represented by the sequence ABACBA. This is a perfectly fitting boundary of very low energy and it could be surprising if it was able

to generate or conduct vacancies, for the formation energy would be almost as high as than in perfect crystal (i.e. ~ 5 ev).

A non-coherent twin boundary is that which separates the twinned from the untwinned material and yet does not contain the twinning vector **b**. Such a boundary (NC) is represented in the diagram (fig. 12) and must contain an appreciably lower density of atoms than the perfect crystal. Thus a boundary of this nature might be expected to conduct vacancies particularly if it lay normal to **b**. As every third (111) plane (the B plane) is common to both the parent and twin, the non-coherent twin boundary will consist of well-defined channels, lying parallel to the (111) twin plane. It should be relatively easy for atoms (and thus effectively vacancies) to flow along these channels, but most difficult for them to flow in any other direction. The boundaries would be expected to conduct vacancies very anisotropically, as appears to be the case. Due to the tie-rod action of every third layer, it would be difficult for extra atoms to be accommodated in this boundary, and it would thus not be a source of vacancies, in the sense we have defined.

The number of possible configurations of a grain boundary is enormous. Some simple boundaries may be considered. A simple tilt boundary between two grains with a small angle of misorientation may be treated as a collection of edge dislocations and a simple twist boundary as a collection of screw dislocations. Such boundaries have low energy and appear to be unable to create vacancies, but may conduct them in the same way as it is thought that dislocation lines may. The conduction of such boundaries should be highly directional. High-energy boundaries have considerable atomic misfit along them and such boundaries should generally be good sources of vacancies. It is a general observation that low energy boundaries are rarely sources of vacancies but high energy ones frequently are.

§ 6. THE EFFECT OF DISPLACEMENT VACANCIES

A large number of vacancies is produced during the bombardment of the target with alpha-particles, and it should be explained why these are not able to permit helium bubbles to form before vacancies from grain boundaries, etc., arrive. One can conclude only that the number of vacancies involved is too small. The following simple model illustrates the situation.

Consider a unit volume of crystal with a simple cubic structure containing no vacancy or interstitial atom sinks which is bombarded with alpha-particles and has attained an equilibrium situation with N vacancies and an equal number of interstitial atoms. Consider also that the temperature is such that the interstitial atoms have a high mobility (ν atomic jumps/sec) and that the vacancies are unable to change their position by thermal activation—this is likely to be the situation in copper at room temperature. For this equilibrium state to be maintained the number of interstitial atoms combining with vacancies in time δt must be balanced by the number ($n\delta t$) being produced in that time. In time δt each interstitial

atom explores $\nu\delta t$ interstitial sites and as each of these enables access to at least four new vacancy positions the number of vacancies annihilated is $4(N/N')\nu\delta t$ where N' is the number of atomic positions in the unit volume. As there are N interstitial atoms, which we shall assume behave independently,

$$4(N/N')\nu\delta tN = n\delta t$$

i.e.

$$N = \frac{1}{2}\sqrt{(nN'/\nu)}.$$

Substituting the values for copper at 600°K bombarded in two hours with alpha-particles sufficient to produce an atomic concentration of 10^{-3} helium atoms in the layer beneath the surface, gives a value for the equilibrium atomic concentration of vacancies to be $\sim 10^{-9}$, i.e. only one vacancy per 10^6 helium atoms. Thus it is reasonable to assume that the annihilation of the vacancies by the interstitial atoms produced at the same time, removes the vast majority of the vacancies. The remainder are so few that they are unable to assist in the formation of the helium bubbles of the size observed. Only if the number of sinks for the interstitial atoms are so numerous that more than one hundredth of the interstitial atoms are annihilated at positions other than at vacancies would these vacancies be expected to affect the observations appreciably.

§ 7. CONCLUSIONS

1. The vacancies introduced during the alpha-particle bombardment are mainly annihilated by the interstitial atoms also produced, so that those remaining are too few to permit the formation of the helium bubbles observed.
2. The necessary vacancies for bubbles to form are drawn from vacancy sources in the material by thermal activation.
3. The nature of a defect boundary in the crystal determines whether it is inherently a source, or merely a conductor of vacancies, or neither.
4. A boundary can be restrained from acting as a source, but not as a conductor, by the application of a lateral pressure.
5. Extremely high pressures are needed to prevent the supply of vacancies.
6. A boundary which conducts vacancies can do so at different rates in different directions.
7. Line defects in the crystal are unable to act as vacancy sources in the conditions of these experiments, but may conduct vacancies.

ACKNOWLEDGMENTS

I am indebted to Mr. D. J. Mazey for his patience and care in obtaining the micrographs, to Dr. J. Coiley and Mrs. B. Jenkins for their help with

the electron microscope, to Mr. G. B. Redding who prepared the targets which were bombarded at Birmingham University under the supervision of Mr. W. Hardy, and to Professor A. H. Cottrell, F.R.S. of Cambridge University for many valuable discussions during the course of this work with which he has been closely associated.

REFERENCES

- BARNES, R. S., 1959, *Colloque sur la Diffusion a l'État Solide* (Amsterdam: North Holland Publishing Co.), p. 57.
BARNES, R. S., and REDDING, G. B., 1959, *J. nucl. Energy, A*, **10**, 32.
BARNES, R. S., REDDING, G. B., and COTTRELL, A. H., 1958, *Phil. Mag.*, **3**, 97.
ELLS, C. E., and PERRYMAN, E., 1959, *J. nucl. Materials*, **1**, 73.
MYKURA, H., 1959, *Phil. Mag.*, **4**, 907.
RICH, J. B., REDDING, G. B., and BARNES, R. S., 1959, *J. nucl. Materials*, **1**, 96.

REVIEWS OF BOOKS

Solid State Physics in Electronics and Communications. Vol. 1, Part 1, Semiconductors. Edited by M. DESIRANT and J. L. MICHELS. (London and New York: Academic Press, 1960.) [Pp. 638.] £6 8s. 6d.

THIS is not a book: it is the Report of a Conference held in Brussels in June 1958. It has all the disadvantages of such a document: (a) eighteen months is far too long a delay: if the papers were worth publishing, they were worth publishing quickly, and three more volumes are yet to appear; (b) many of the papers are mere interim accounts of work in progress; they may have been worth presenting verbally at the conference, but are not worth printing; (c) there is no reason to suppose that any of these papers were refereed; they do not, therefore, carry the weight of work published in reputable scientific journals; (d) it is more lavishly and expensively printed than such ephemeral work deserves; no private person would afford to buy it for the occasional reference; (e) although libraries may be persuaded into purchasing it, the problem of finding papers in it by reference will be much more complicated than in the ordinary scientific journals; even the title is too long for a neat footnote, and who can guess how it will be indexed in the book stacks.

For these reasons I cannot commend the publication of such works, however admirable may be some of the actual research that is here reported (and some of it is quite interesting). Perhaps the publishers or the organizers of the conference can explain the rationale of their procedure, and tell us its advantages compared with the publication of ordinary scientific papers through the usual channels in the recognized journals.

J. M. Z.

A New Method in the Theory of Superconductivity. By N. N. BOGOLIUBOV, V. V. TOLMACHEV and D. V. SHIRKOV. (New York: Consultants Bureau, Inc.; London: Chapman & Hall, Ltd.) [Pp. 121.] Price \$5.75, about £2 10s.

THE theory of superconductivity has undergone a rapid development in the last three years. Indeed it would be fair to say that physicists now have a qualitative understanding of many of the main features of the superconducting state. The initial impetus for this development came from the now celebrated paper of Bardeen, Cooper and Schrieffer. Immediately after this Bogoliubov, Tolmachev and Shirkov developed an alternative approach to the theory. The volume under review is essentially a translation of this and several other Russian papers on the subject. As such it will be of the first importance to those theoretical physicists whose Russian is inadequate to deal with the original papers. It is stated on the jacket that the book contains a complete solution to the theoretical problems of superconductivity. This, at least in the opinion of the reviewer, is not so and it is unlikely that the authors themselves would support such a claim. The book does, however, contain a clear account of the Russian contributions to the subject. The main criticism that might be made of it is that the presentation is rather abstract and that very little physical discussion of the theoretical results is given. Nevertheless the serious student of the subject will find it particularly valuable for the clear account of the quasi particle transformation and the principle of compensation of dangerous graphs, for the discussion of the role of the coulomb interaction and for the detailed investigation of the collective modes of motion in the superconducting state. There are unfortunately several obscurities in the translation, rather too many considering the price of the book.

G. V. C.

The Conceptual Foundations of the Statistical Approach in Mechanics. By P. and T. EHRENFEST. Translated by M. J. Moravcsik. (Ithaca, N.Y.: Cornell University Press.) [Pp. xii+114.] Price £1 4s.

THIS is a most welcome translation of one of the few classics in the field of statistical mechanics. It has been relatively difficult to get hold of the original (at any rate until the appearance last year of the Collected Works of P. Ehrenfest) and, as is pointed out by Kac and Uhlenbeck in a short preface, the original is couched in rather difficult German. Anybody who is interested in the development of statistical mechanics, and especially in its foundations, will want to possess this booklet, and luckily its price does not have to be a hindrance, as it so often is these days. It is rather unfortunate that the translation is far from smooth and reads even less easily than the original. A number of German terms were deemed to be untranslatable, but I fail to see why recurrence paradox, reversibility paradox, and equivalent would be incorrect renderings of *Wiederkehrerwand*, *Umkehrerwand* and *gleichberechtigt*, to name a few instances. A personal element is introduced by a special preface for this translation (also printed in the Collected Works) by Mrs. Ehrenfest consisting of a few remarks which were omitted in the original on her instigation, but which she feels should now be included.

D. TER HAAR.

BOOK NOTICES

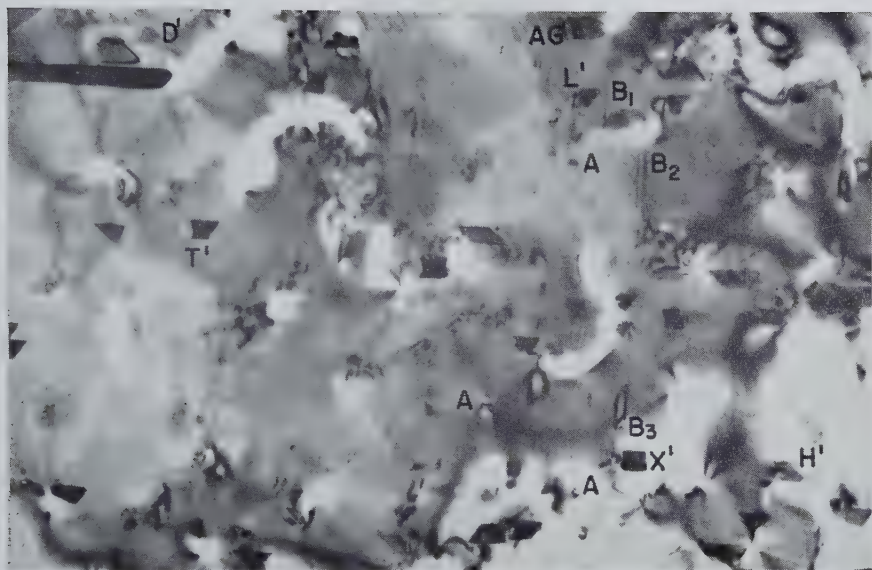
Confluent Hypergeometric Functions. By L. J. SLATER. (Cambridge: University Press, 1960.) [Pp. 247.] Price £3 5s.

A Simple Approach to Electronic Computers. By E. H. W. HERSEE. (London: Blackie & Son.) [Pp. 104.] Price 12s. 6d.

Lectures on Fourier Integrals. By S. BOCHNER. Translated by M. Tenenbaum and H. Pollard. *Annals of Mathematics Studies*, No. 42. (Princeton: University Press; London: Oxford University Press.) Price £2.

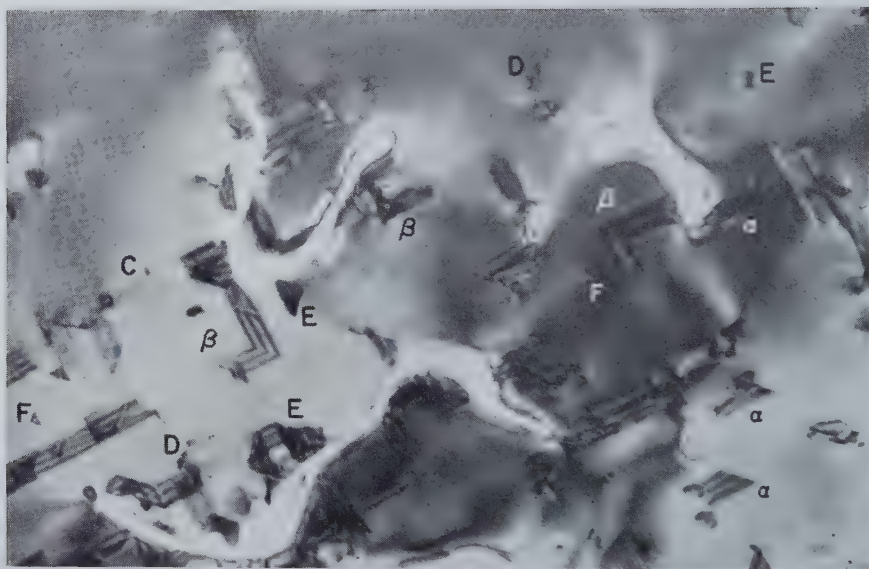
[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

Fig. 2



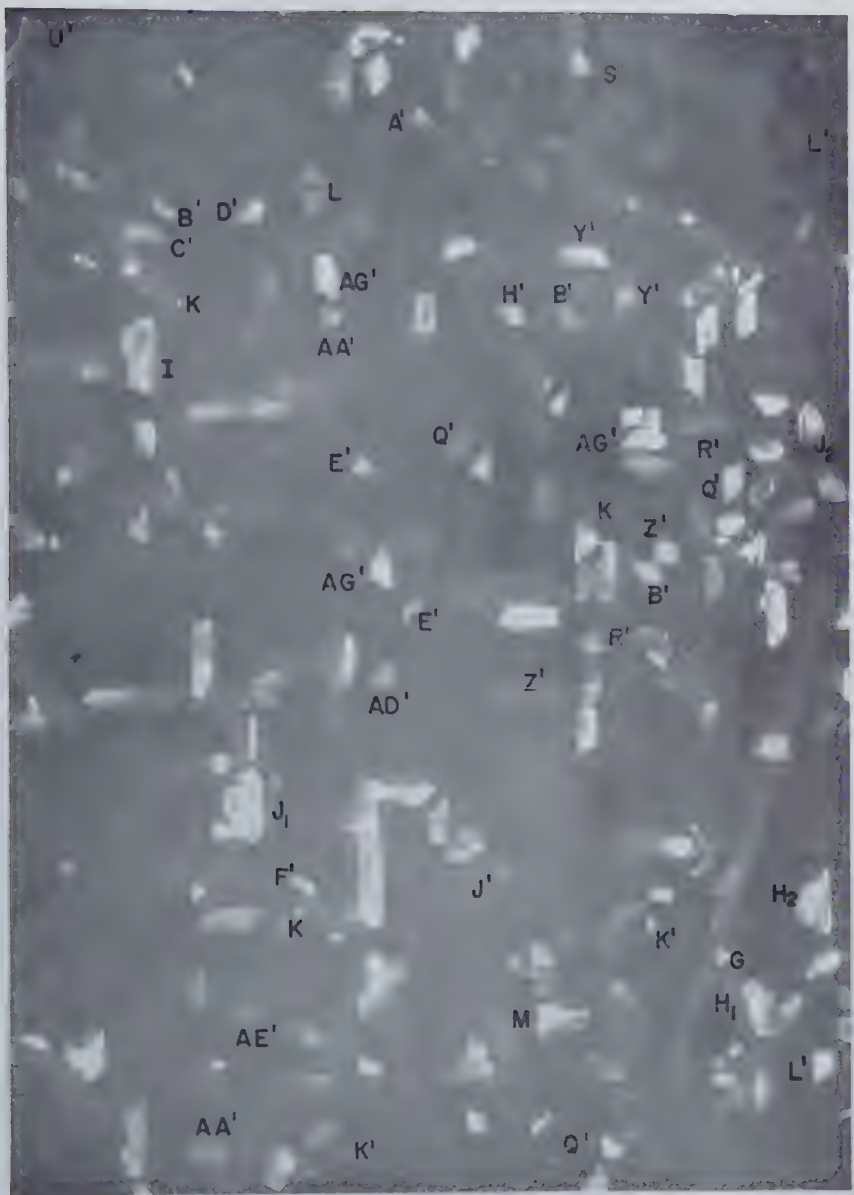
Structure of fresh area of monocrystalline film of silver made by evaporation onto heated rock salt. Note simple stacking faults AG', D', X', etc., dislocation loops A, and parallel line features B₁₋₃. Holes appear in white contrast. Foil thickness about 650 Å. $\times 60\,000$.

Fig. 3

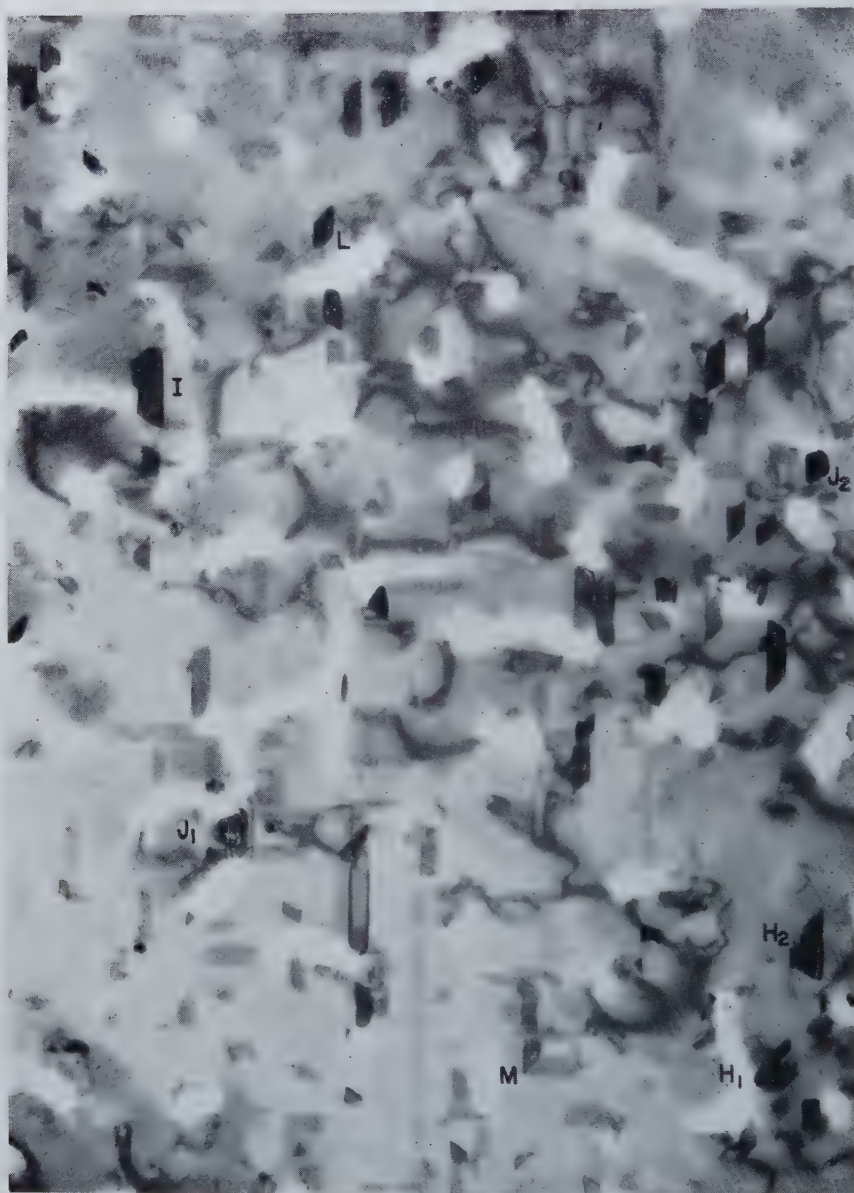


Region showing line or point defects C, whole dislocations D, stacking faults α , β , stacking-fault tetrahedra E, and triangular loops F which are probably Frank sessile dislocations. The simple stacking faults α and complex faults β show typical extinction bands. Foil thickness 500–650 Å. $\times 100\,000$.

Fig. 4 (a)

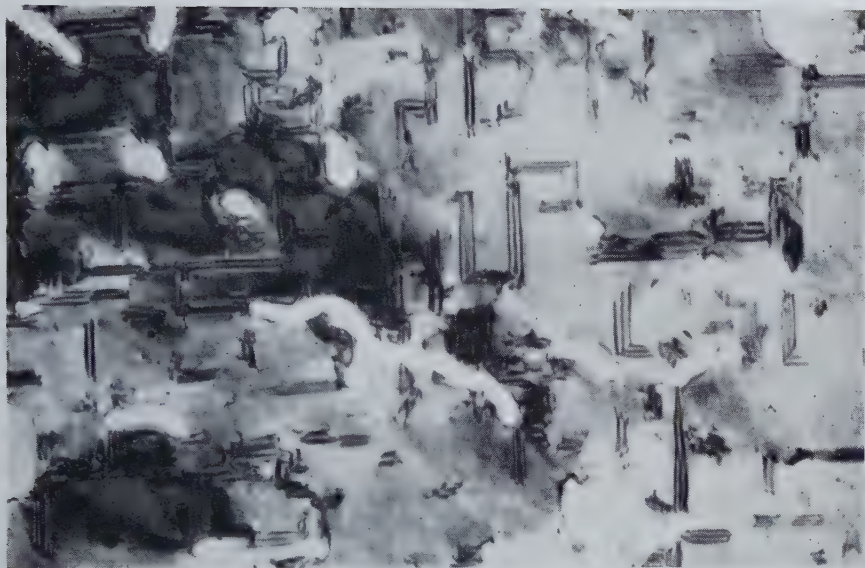


(a) Dark field electron micrograph of fresh area of film showing stacking faults in white contrast. Feature I is probably a twin. The interpretation of features J₁ and J₂ is in doubt (see text). The remaining features are stacking faults and may be classified according to shape as described in the text. Note groups of parallel faults H₁ and H₂. ×60 000.

Fig. 4 (b) (*continued*)

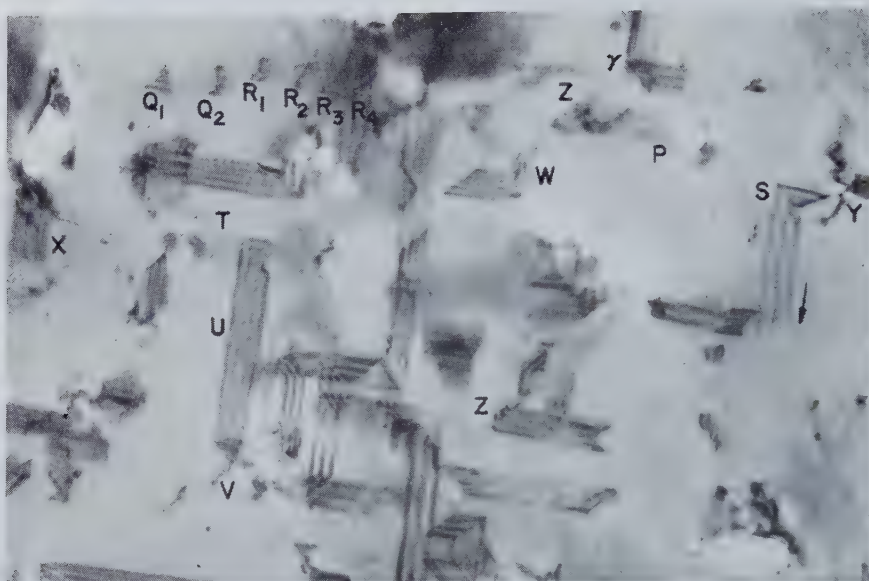
- (b) Light field micrograph of same area as (a) taken immediately following. In the time interval between exposures, pairs of simple faults at L and M have interacted and become single-bend faults. Faults now appear in dark contrast or show finely spaced extinction bands. $\times 60\,000$.

Fig. 5



Typical appearance of film after long thermal stressing in the beam. Contrast figs. 2 and 4. The faults here have widened and interacted to produce complex faults. Holes, dislocations and other faults act as barriers to further dissociation. Holes and thin regions act as pinning points for partial and stair-rod dislocations. $\times 60\,000$.

Fig. 6



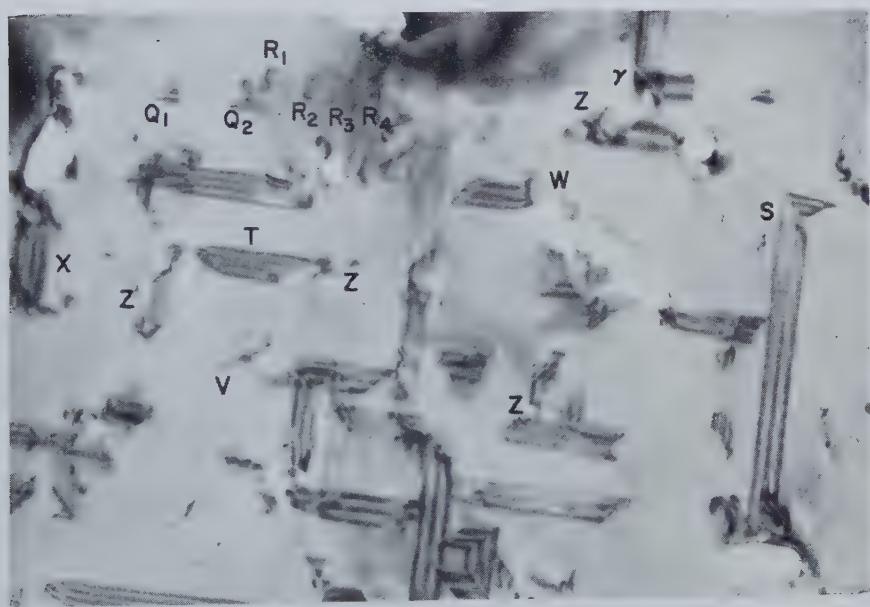
(a)

Sequence of micrographs showing effect of thermal stressing on structure. Note pile-up of dislocations Q_1 , Q_2 , R_1 – R_4 against a group of stacking faults; dissociation is least at the head of the pile-up. Q_1 and Q_2 are on a different but probably parallel $\{111\}$ plane to that of R_1 – R_4 . The hole edge W appears to be a source of dislocations, since a second fault on a different plane has appeared between exposures (a) and (b) forming a single-bend fault, while in (c) a third fault on a $\{111\}$ plane parallel to that occupied by the fault in (a) has appeared. The single-bend fault S has widened progressively between (a) and (b) and between (b) and (c), being blocked by a hole in (c). Complex fault UT has changed to a simple fault T in (c) permitting dislocation V in (b) to move to a new position in (c). There appear to be two parallel faults at X which have extended between (b) and (c). Dislocations at Y and Z in (a) are held up at faults. Z has moved away in (c). Note the pinning of partials P and W at holes, in (a). Note the tendency of partials associated with faults Q_1 , Q_2 and W in (c) to be parallel to $\langle 110 \rangle$ directions in the fault plane. Stair-rod dislocations as at the bend in fault S in (a) are parallel to $\langle 110 \rangle$ directions in the fault plane. $\times 140\,000$.

Fig. 6 (*continued*)

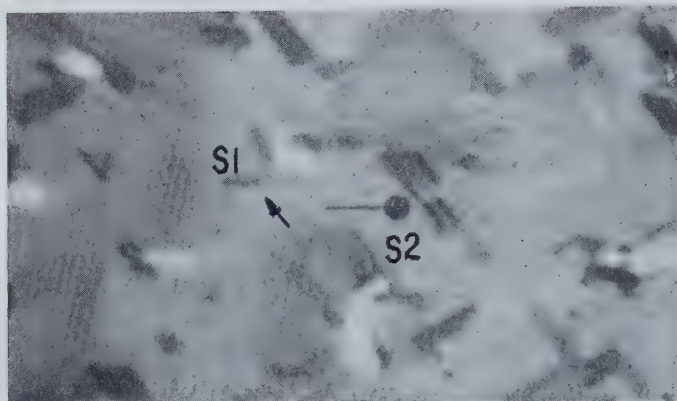


(b)



(c)

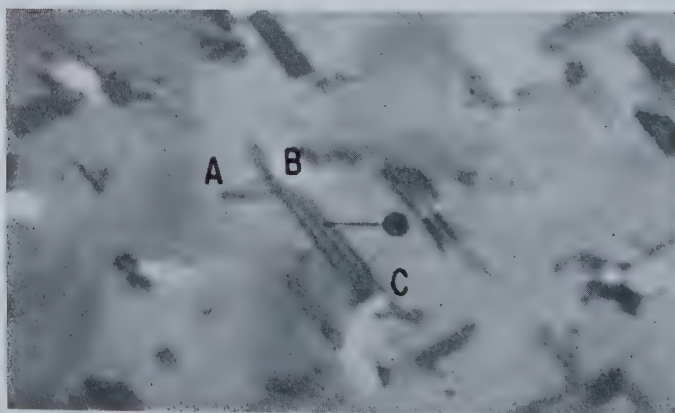
Fig. 7



(a)



(b)



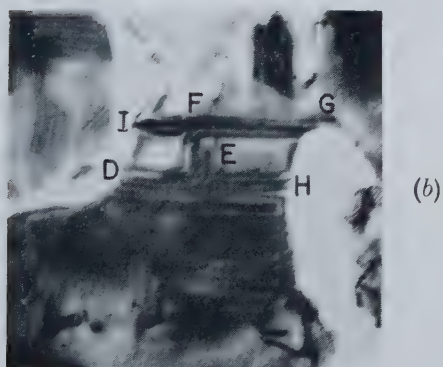
(c)

Sequence of ciné photographs showing combination of two simple faults to form a single-bend complex fault. In (a) the two faults S1 and S2 are visible. In (b) fault S2 has extended but the two partials (retouched for reproduction) indicated by arrows are separately visible. In (c) the two partials have combined giving a stair-rod dislocation at B (not visible) and forming a single-bend fault ABC (angle between faulted planes acute). Magnification about $\times 80\,000$ parallel to screen marker and about $\times 67\,000$ perpendicular to this.

Fig. 8



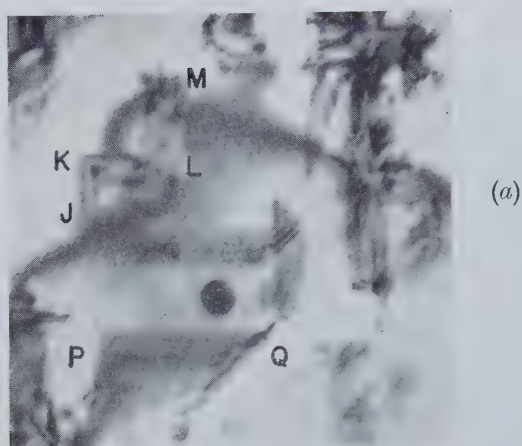
(a)



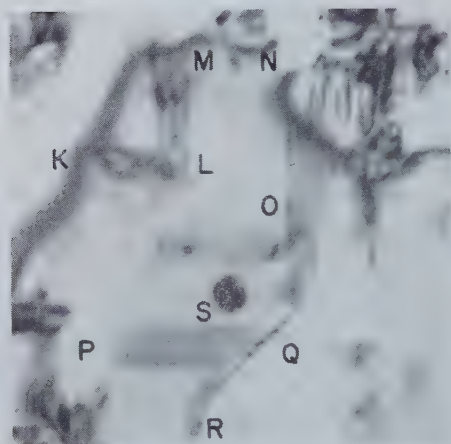
(b)

Sequence of micrographs showing conversion of single-bend fault DEF in (a) to a complex double-bend fault DEHGFI in (b). $\times 140\,000$.

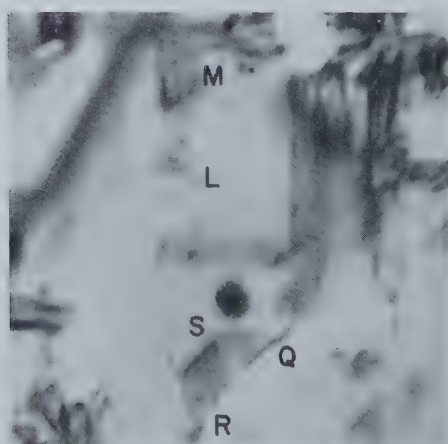
Fig. 9



(a)



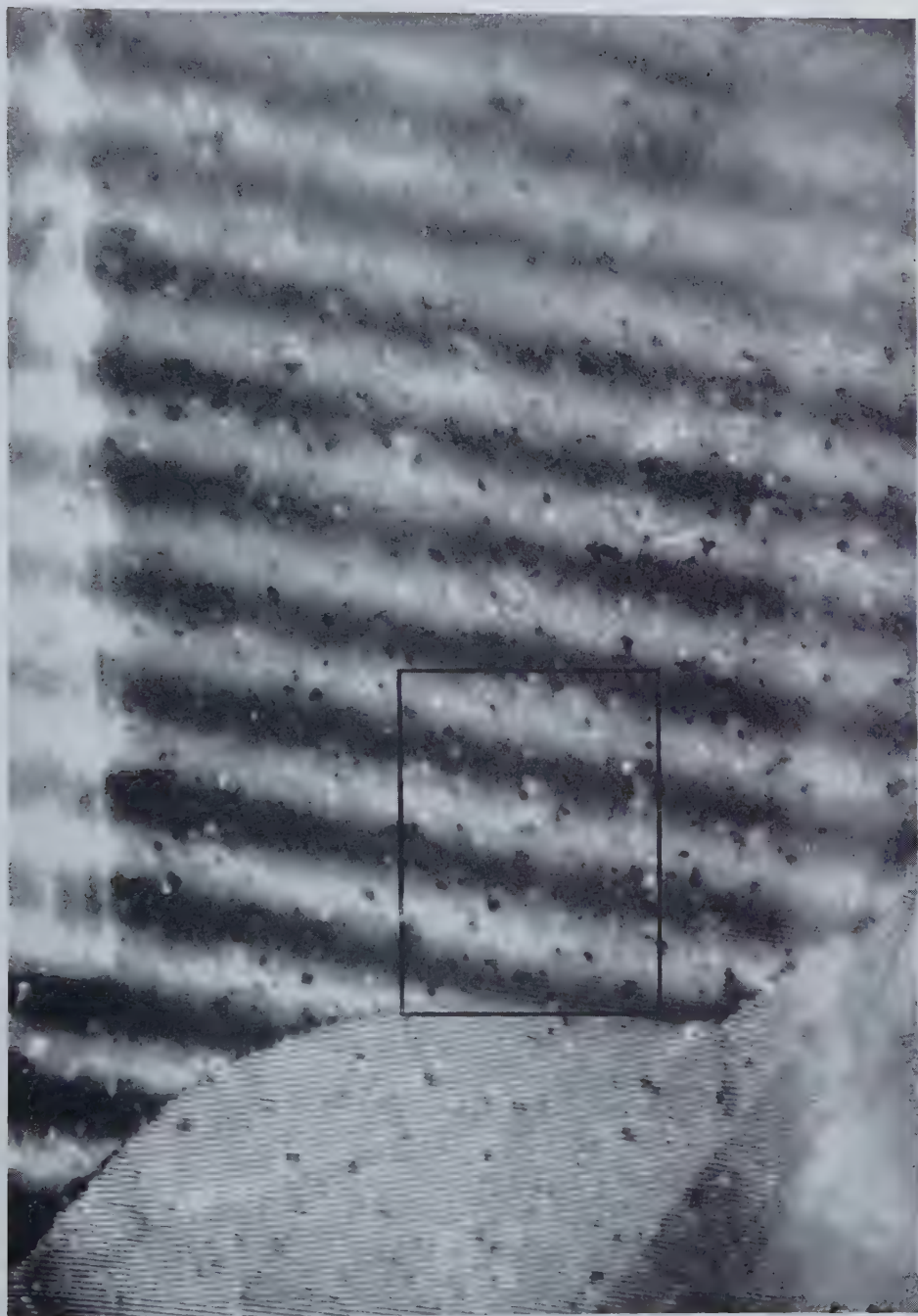
(b)



(c)

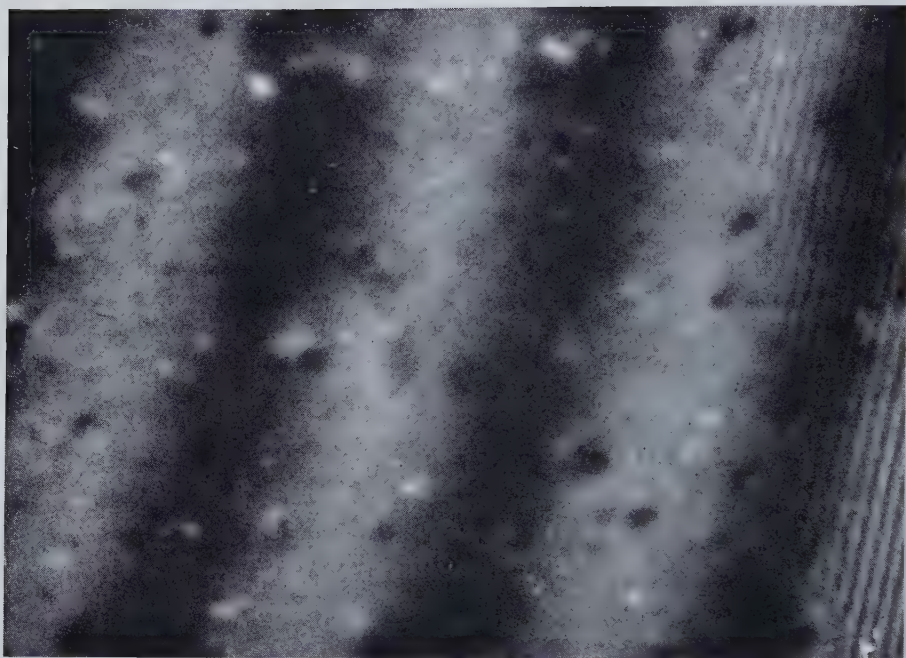
Sequence of micrographs showing changes in fault configuration. Between exposures (a) and (b): the double-bend-fault JKLM has become a single-bend fault KLM, a simple fault NO has formed between a hole at N and another complex fault at O, and a simple fault RS has formed between a hole at R and a fault PQ. Between exposures (b) and (c): the single-bend fault KLM has become a simple fault ML, fault PSQ has contracted to SQ and its partial appears to be held up at the partials of fault SR. $\times 140\,000$.

Fig. 1



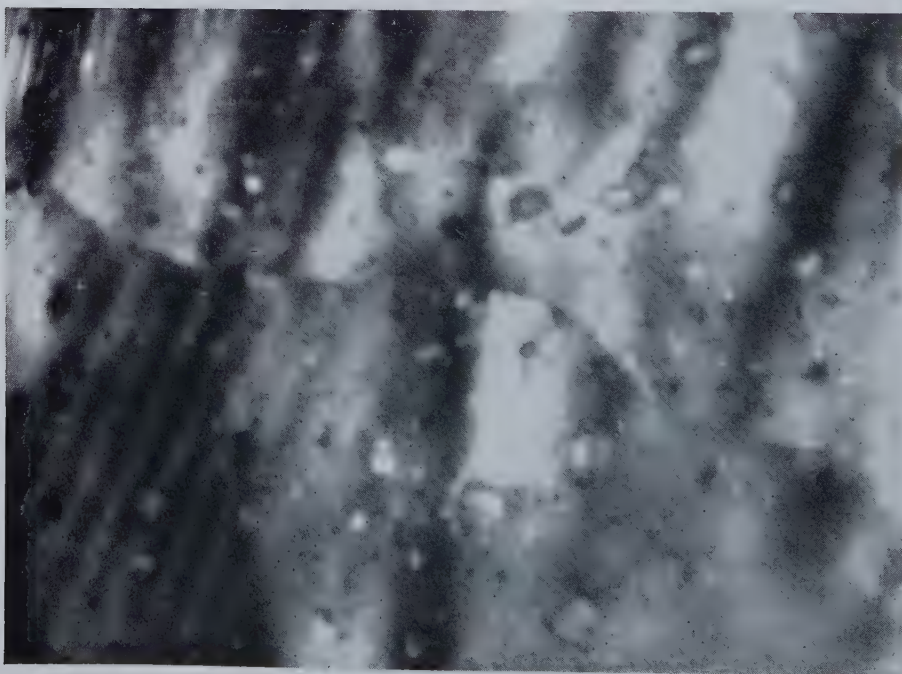
Graphite irradiated by 10^{20} n/cm² at 30°C, 320 000 \times .

Fig. 2



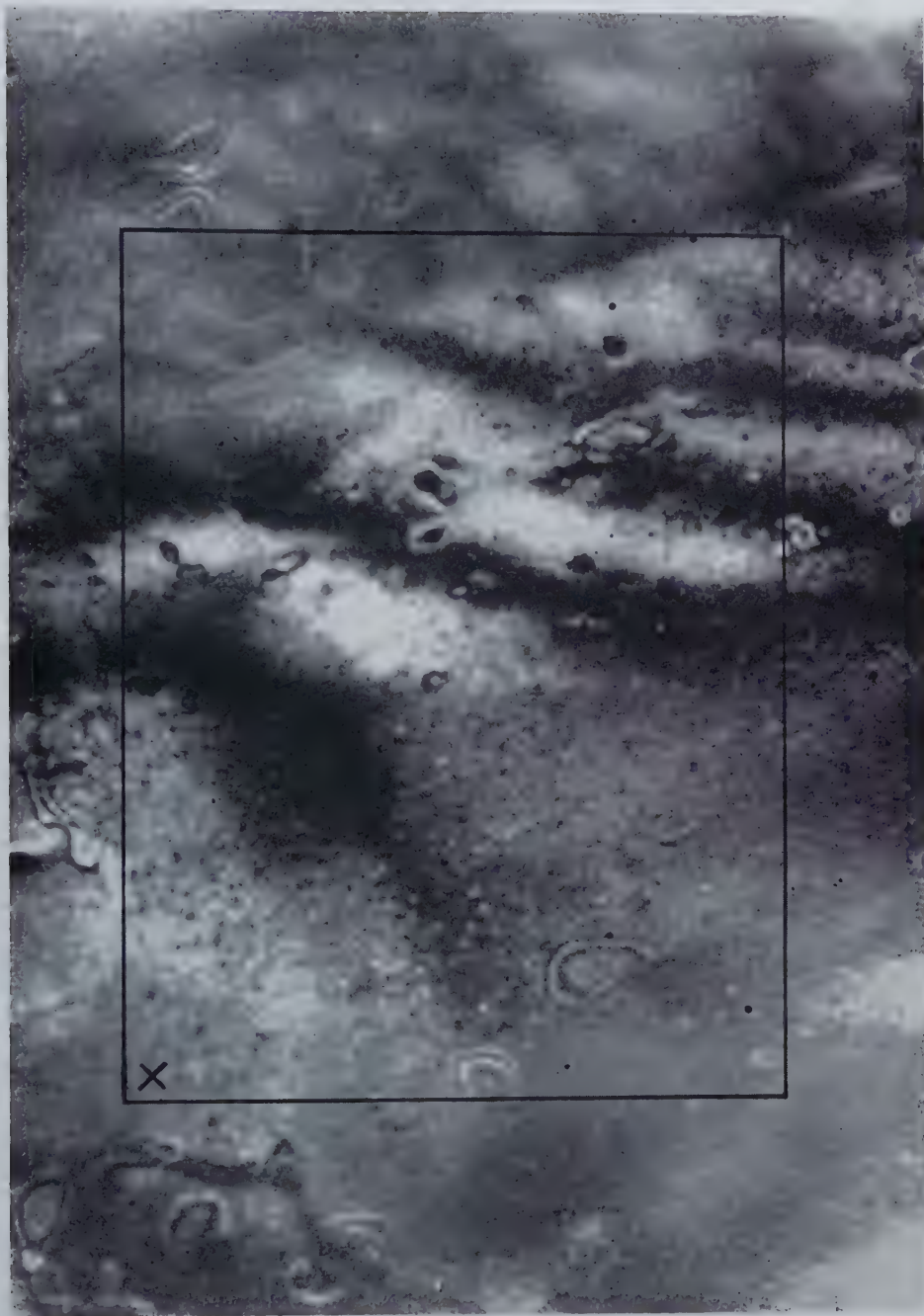
Detail of fig. 1, $800\,000\times$.

Fig. 3



Same type of graphite as fig. 1 showing loops, $320\,000\times$.

Fig. 4



Same graphite as fig. 1 annealed at 400°C , $320\,000\times$.

Fig. 5



Selected area of the specimen fig. 4 in bright-field illumination. $320\,000\times$.

Fig. 2



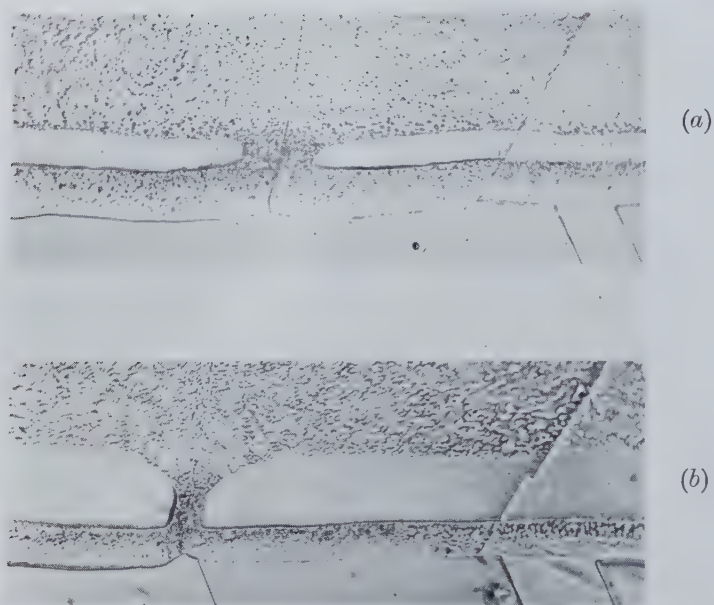
The incomplete precipitation of injected helium atoms in a single crystal of Johnson Matthey spectroscopically standardized copper during heating at 700°C for 1 hour. $\times 570$.

Fig. 3



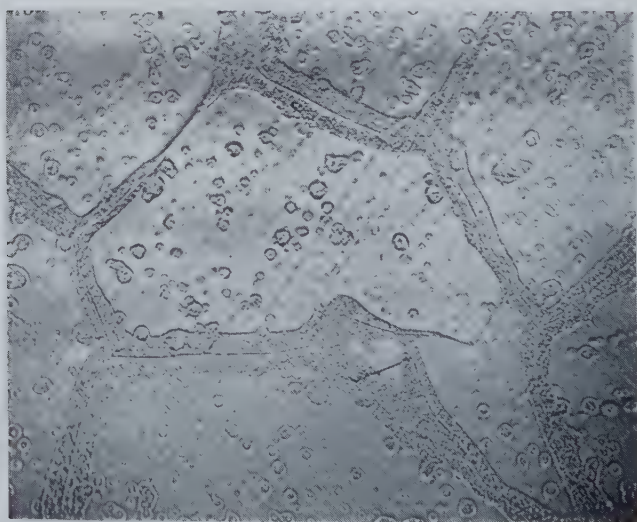
Bubbles at a free surface in copper after heating to 800°C for 1 hour.
 $\times 750$.

Fig. 4



The surface exposed during heat treatment of 1 hour at 825°C (a) after a thin layer has been removed by mechanical and electro-polishing and (b) after the similar removal of a further layer. $\times 750$.

Fig. 5



A parallel section of a similar piece of fine grain copper after an anneal of 1 hour at 750°C . $\times 570$.

Fig. 6.

A parallel section showing bubble formation at twin boundaries. $\times 570$.

Fig. 7

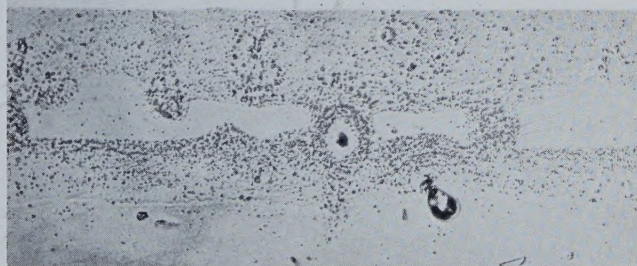
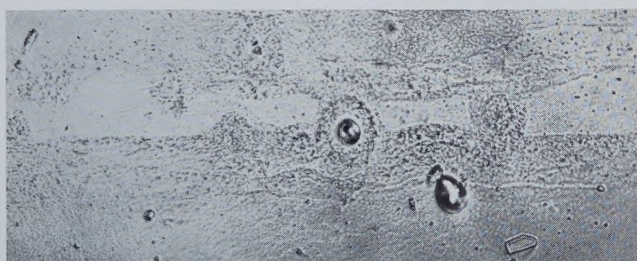
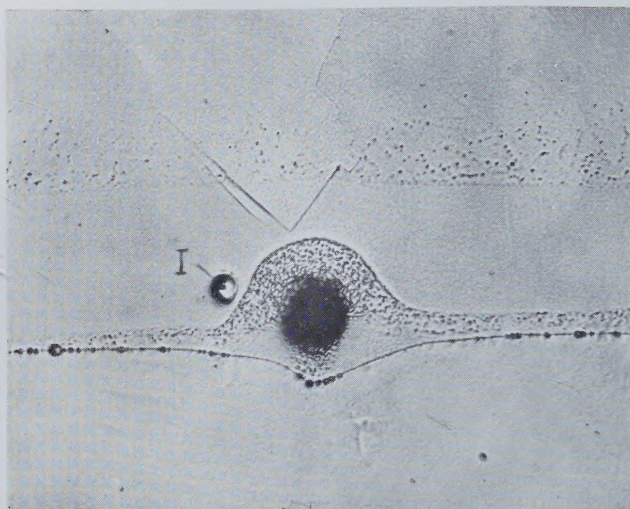
Microphotographs of a void in Johnson Matthey spectroscopically standardized aluminium (a) a section almost through the centre and (b) one below the centre. $\times 570$.

Fig. 8



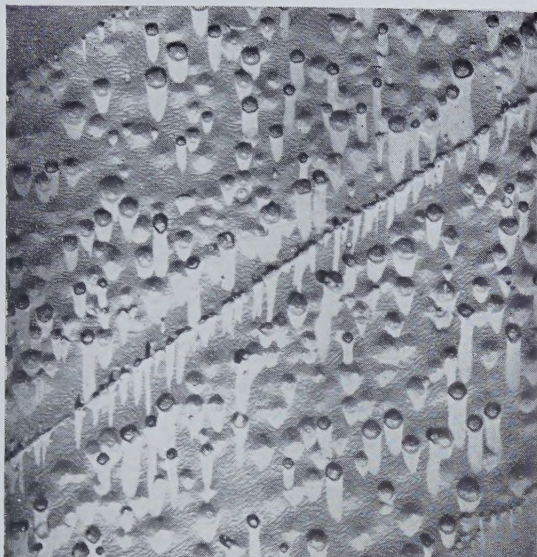
Photomicrograph of an inclusion (I) and a void in copper showing their relative ability to supply vacancies. $\times 1000$.

Fig. 9

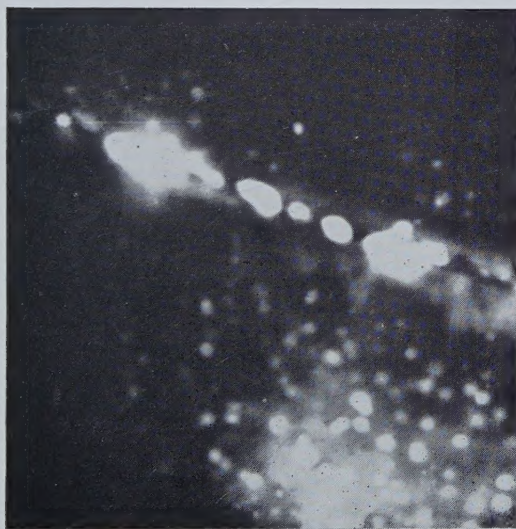


Electron micrographs of a circular region of bubbles in aluminium. $\times 9000$.

Fig. 10



(a)



(b)

Electron micrograph of a grain boundary showing a zone with no bubbles adjacent to the boundary (a) shadowed carbon two-stage replica $\times 7000$ and (b) transmission through a thin film of the metal extracted from the helium layer. $\times 12000$.

Fig. 11



A grain boundary which is either a good vacancy source (central region) or a poor conductor, depending upon whether the adjoining grains are twinned or not. $\times 3000$.

Prof. H. W. -